An exploratory study of factors related to effectiveness of Free/Libre Open Source Software teams

Kevin Crowston and U. Yeliz Eseryel
Syracuse University School of Information Studies
crowston@syr.edu, yeliz2002@alum.syracuse.edu

http://floss.syr.edu/

Our* research question

What work practices make some FLOSS teams more effective than others?

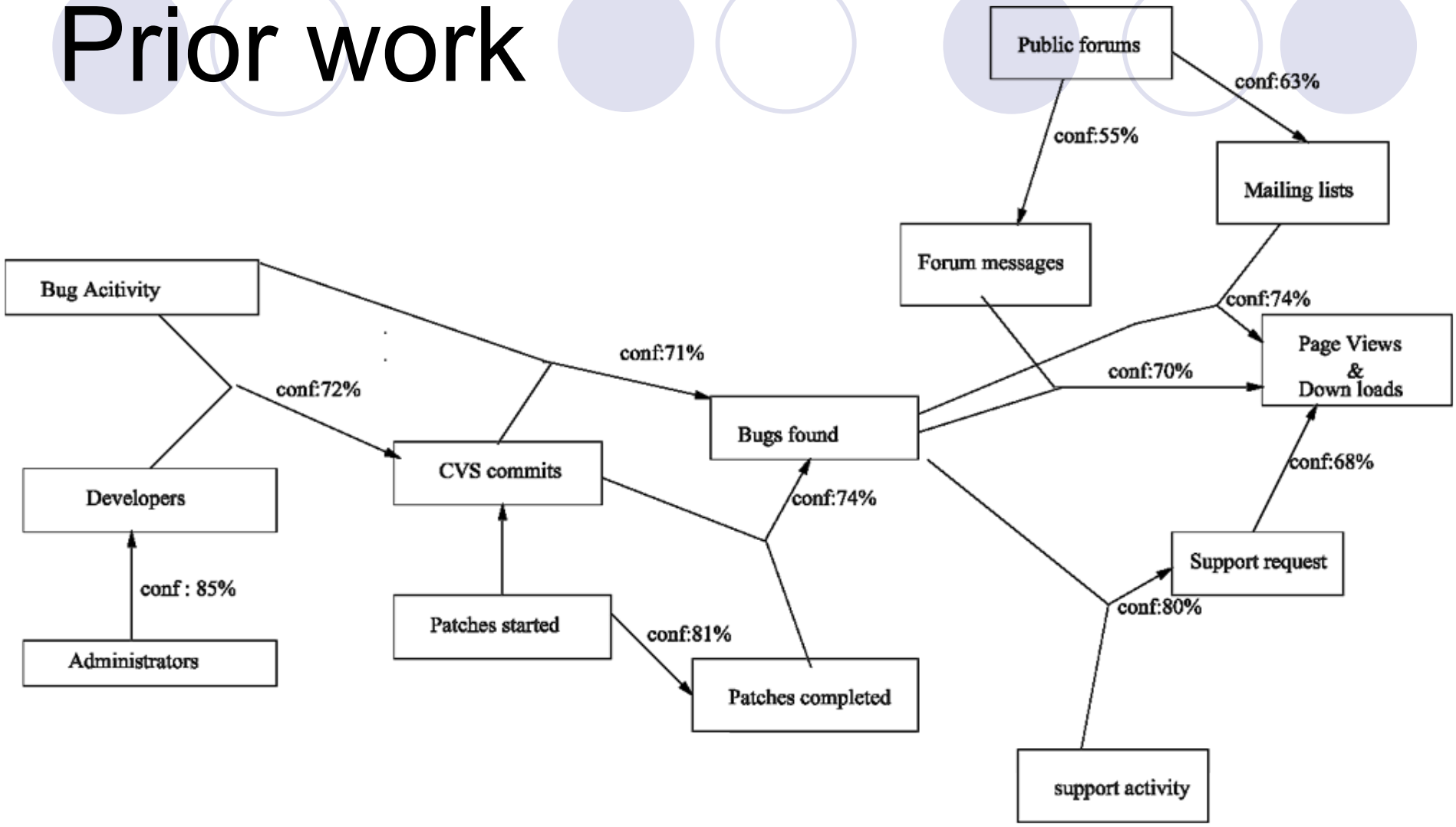
* We = Kevin Crowston, Hala Annabi, James Howison, Chengetai Masango, Yeliz Eseryel, Kangning Wei and Qing Li

Partially supported by US NSF Grants 03-41475 and 04-14468.

Question for this paper

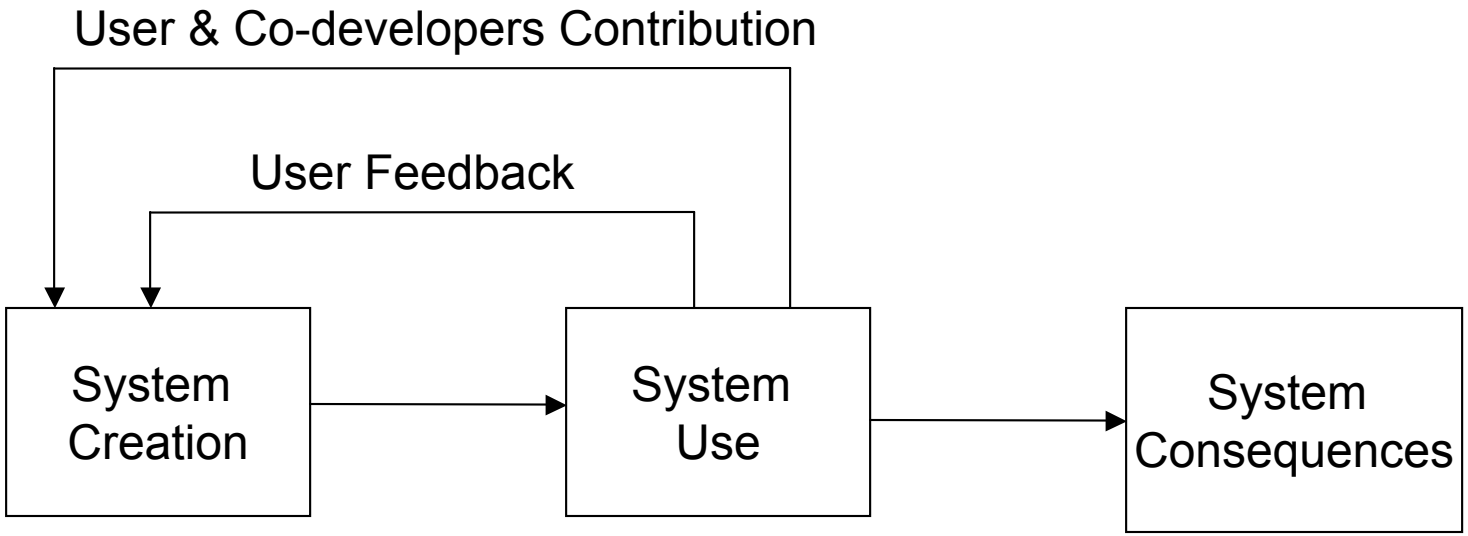
- What variables about the projects are associated with effectiveness?
- What relationships exist among these variables (suggesting which practices are most important)?
- Which trends can we observe that are useful for further study?
- Which projects would be most informative for follow up study?

Prior work



Source: Chawla, S., Arunasalam, B., & Davis, J. (2003). Mining open source software (OSS) data using association rule network (Technical Report No. 535). Sydney: School of Information Technologies, University of Sydney.

Our success model



- Progress of process
- Number and frequency of releases
- Number of developers

- "Popularity"
  - Downloads
  - Page views

- Developer satisfaction

Data

SourceForge.net Statistics

Registered Projects: **98,294**
Registered Users: **1,046,676**

Project: Compiere ERP + CRM Business Solution: Summary

Smart ERP+CRM solution for Small-Medium Enterprises in the global market covering all areas from order and customer/supplier management, supply chain to accounting. For $5-500M revenue companies looking for "brick and click" first tier functionality.


Project of the month for: February 2004

- Database Environment: JDBC, Project is a database abstraction layer (API), Oracle, Sybase
- Development Status: 5 - Production/Stable
- Intended Audience: Developers, End Users/Desktop, Customer Service, Financial and Insurance Industry, Information Technology
- License: Mozilla Public License 1.1 (MPL 1.1)
- Operating System: 32-bit MS Windows (NT/2000/XP), All 32-bit MS Windows (95/98/NT/2000/XP), All POSIX (Linux/BSD/UNIX-like OSes), OS Independent (Written in an interpreted language), Linux, Solaris, Win2K, WinXP, Microsoft Windows Server 2003
- Programming Language: Java, JavaScript, PL/SQL
- Topic: Dynamic Content, CRM, ERP, Accounting, Point-Of-Sale, Build Tools
- Translations: Chinese (Traditional), English, French, German, Italian, Portuguese, Spanish
- User Interface: Gnome, KDE, Win32 (MS Windows), Web-based

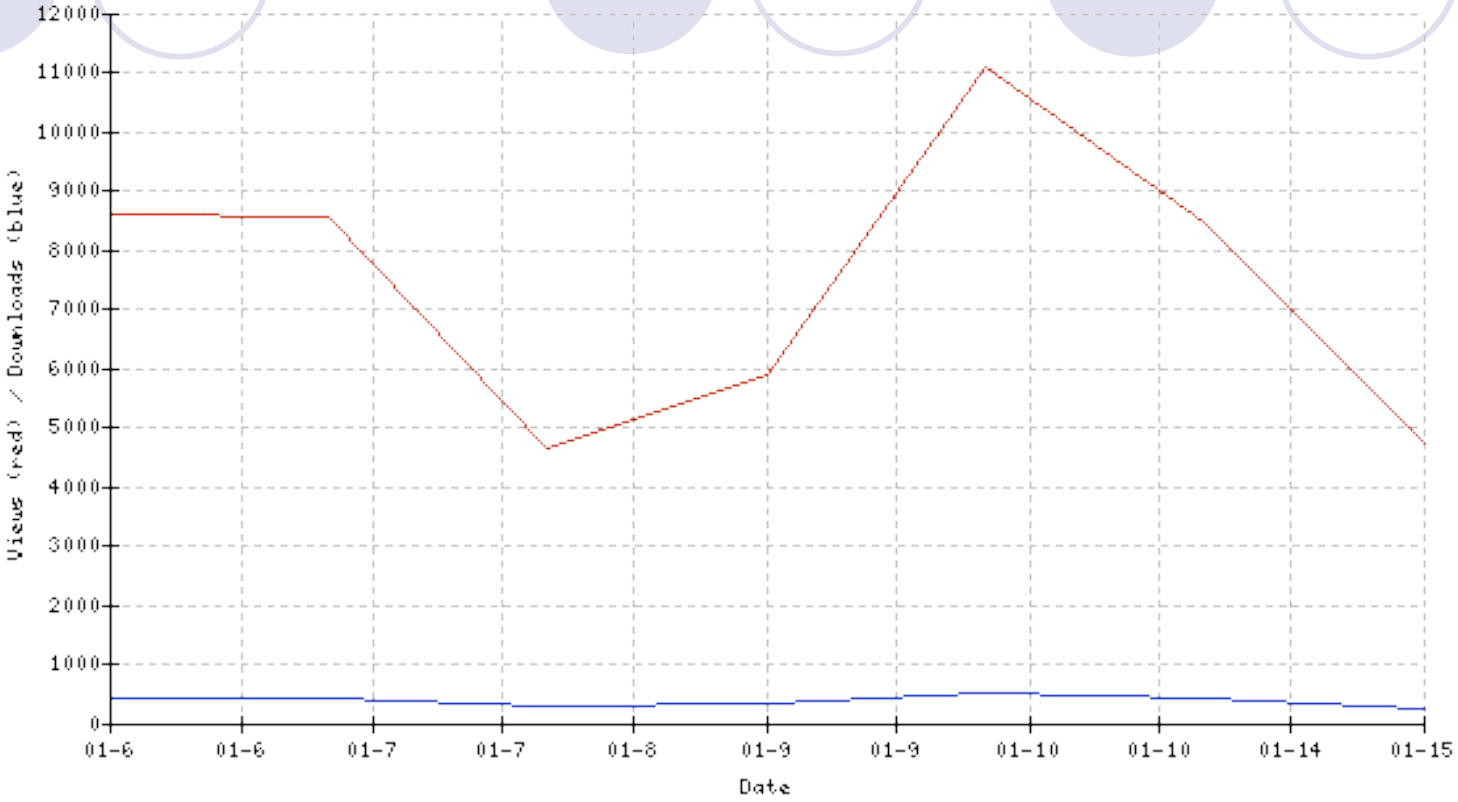
Developer Info

Project Admins:
jjanke
kmpink
sklakken

Developers: 41
[View Members]

Usage Statistics

SourceForge.net Statistics: Compiere ERP + CRM Business Solution



Statistics for the past 7 days

Date	Rank	Page Views	D/l	Bugs	Support	Patches	All Trakr	Tasks	CVS
15 Jan 2005	12 (99.93)	4,746	258	0 (0)	1 (0)	0 (0)	1 (0)	0 (0)	0
14 Jan 2005	11 (99.94)	8,469	414	1 (0)	6 (9)	0 (0)	7 (9)	0 (0)	0
10 Jan 2005	6 (99.97)	11,081	491	2 (1)	6 (6)	0 (0)	8 (7)	0 (0)	0
9 Jan 2005	4 (99.98)	5,893	340	0 (0)	0 (2)	0 (0)	0 (2)	0 (0)	0
8 Jan 2005	3 (99.99)	4,641	267	0 (0)	1 (1)	0 (0)	1 (1)	0 (0)	0
7 Jan 2005	4 (99.98)	8,563	402	1 (3)	6 (7)	0 (0)	7 (10)	0 (0)	0

Data

- Spidered SourceForge website in January 2001 and April 2003
  - Raw data available through OSSMole
- Limited analysis to projects that had had downloads
  - 4512 projects in total

Summary of data

- Topic (11 categories)
- Rank percentile, Page Views, Downloads, CVS activity
  - in 2003, in 2001, change from 2001 to 2003
- Bugs, Support requests, Patches, Tasks
  - in 2003, percentage closed in 2003, in 2001, percentage closed in 2001, change 2001 to 2003
- Administrators, Developers
  - in 2003
- Age

Analysis approaches

- Factor analysis
  - To understand the relation among variables
- Regression
  - To understand the key factors that explain "FLOSS popularity" components:
    - Downloads
    - Page Views
- Association rule data mining
  - To find key rules between our important variables identified in regression analysis

Factor Analysis

	Component								
	1	2	3	4	5	6	7	8	9
Downloads_d	.888								
Downloads0302	.884								
Page_Views0302	.811								
Rank_p0302	.706								
Page_Views_d	.616								
Rank_d		-.871							
Rank_p0101		.846							
Age0101		.716						.432	
Downloads0101	.443	.697							
Page_Views0101	.433	.630							
Patches0302			.783						
Patches0101			.767						
Patches_clp0101			.742						
Patches_clp0302			.688						
CVS0302				.868					
CVS_d	.328			.804					
CVS0101		.478		.656					
Bugs_clp0101					.825				
Bugs_clp0302					.737				
Bugs0101					.641				
Bugs0302	.436		.305		.563				
Support_clp0101						.739			
Support0101						.714	.312		
Support0302	.333					.694	.321		
Support_clp0302						.656			
Admins0302							.774		
Developers0302							.699		
Tasks0101								.866	
Tasks_clp0101								.844	
Tasks0302									.859
Tasks_clp0302									.800

Popularity 2003

Popularity 2001 (and age)

Patches

CVS

Bugs

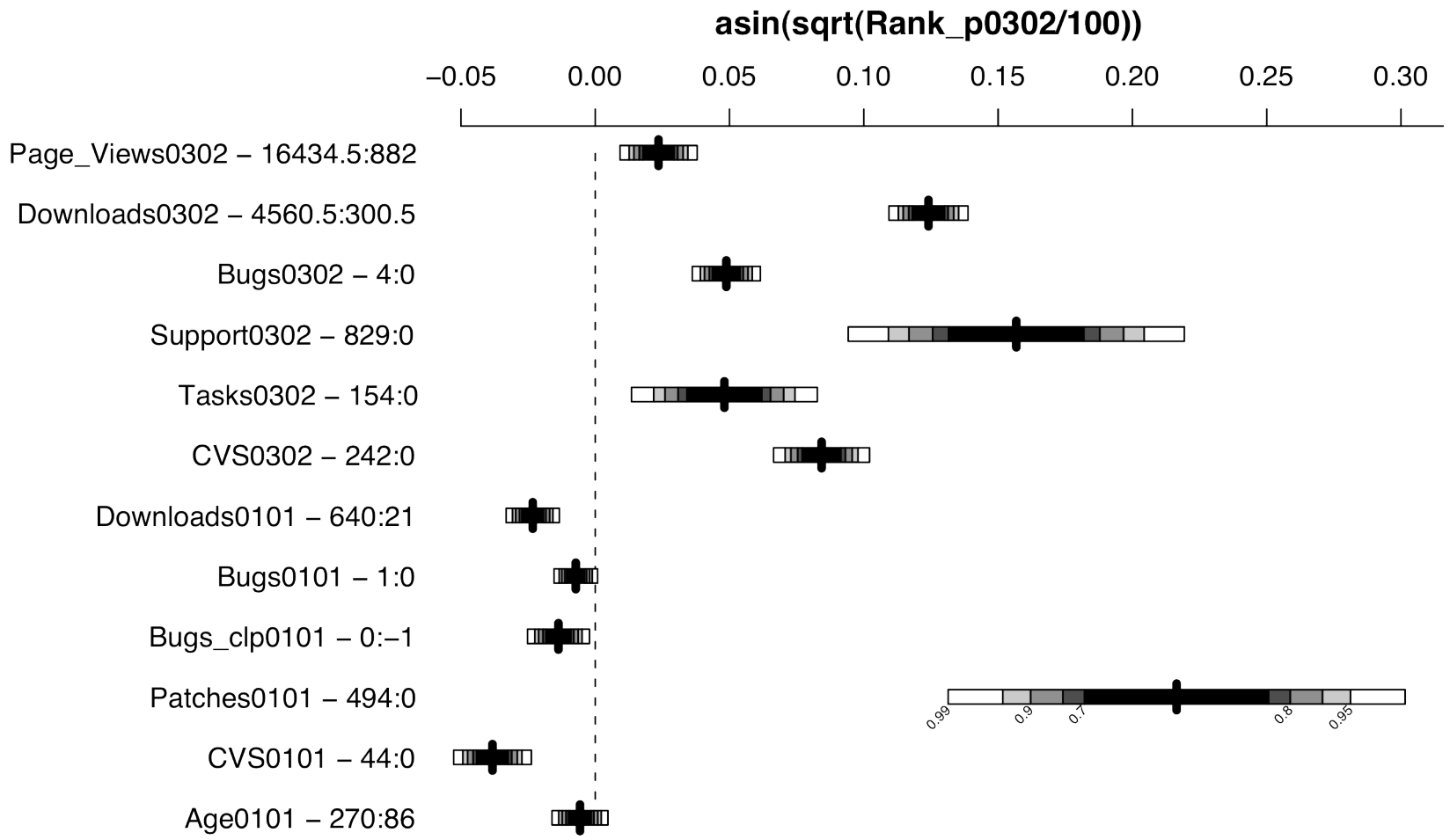
Support

Developers

Tasks 2001

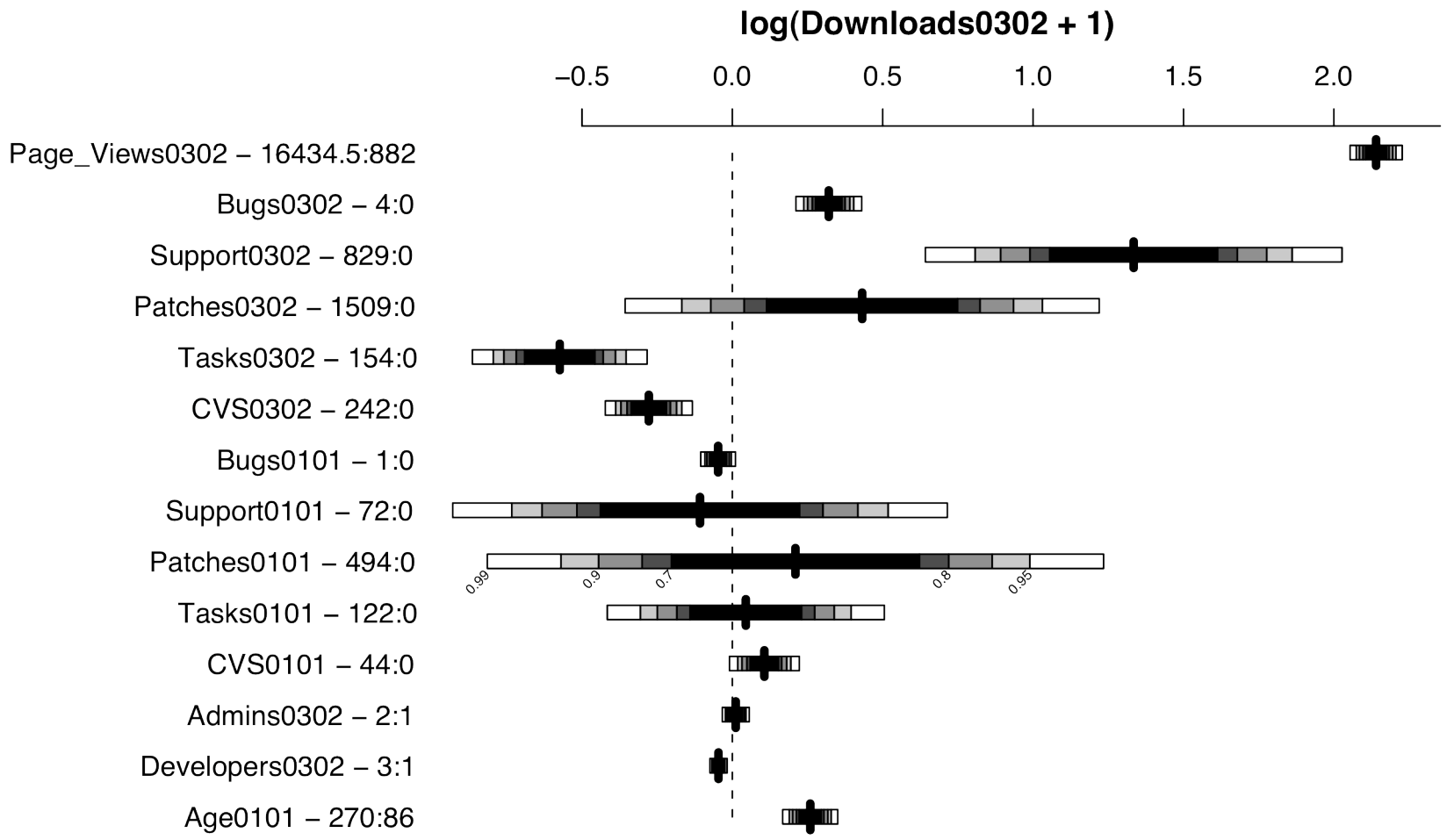
Tasks 2003

Regression Analysis Results (1)



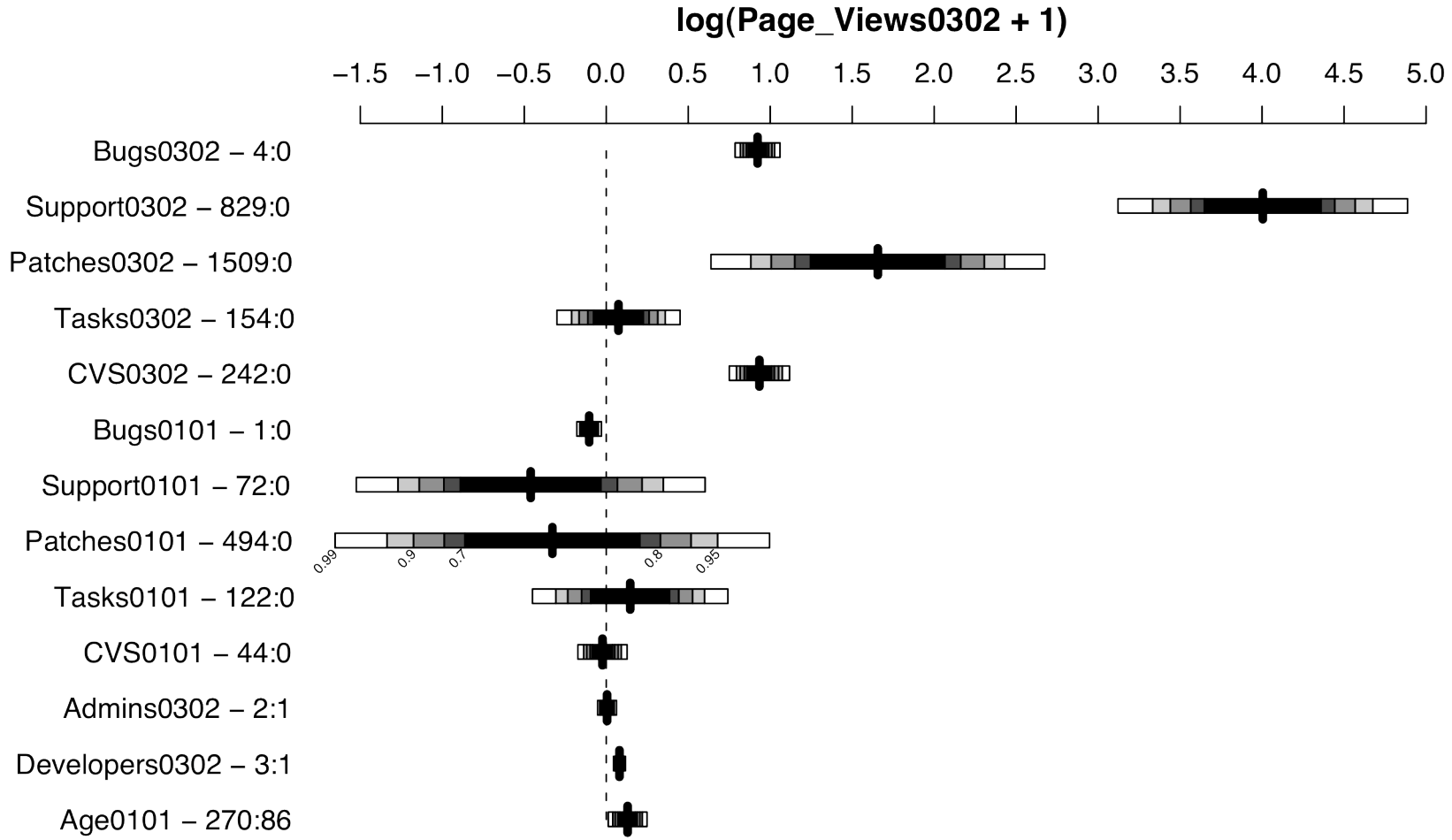
R^2=0.55

Regression Analysis Results (2)



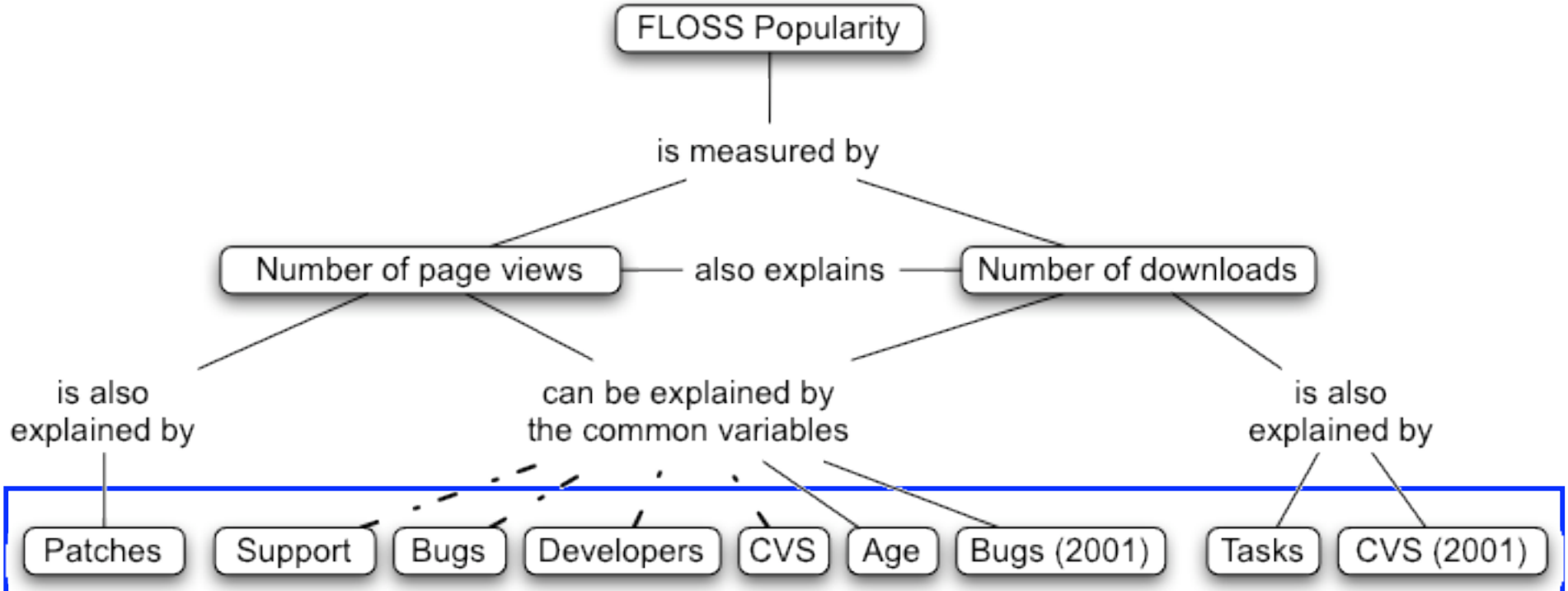
R^2=0.69

Regression Analysis Results (3)



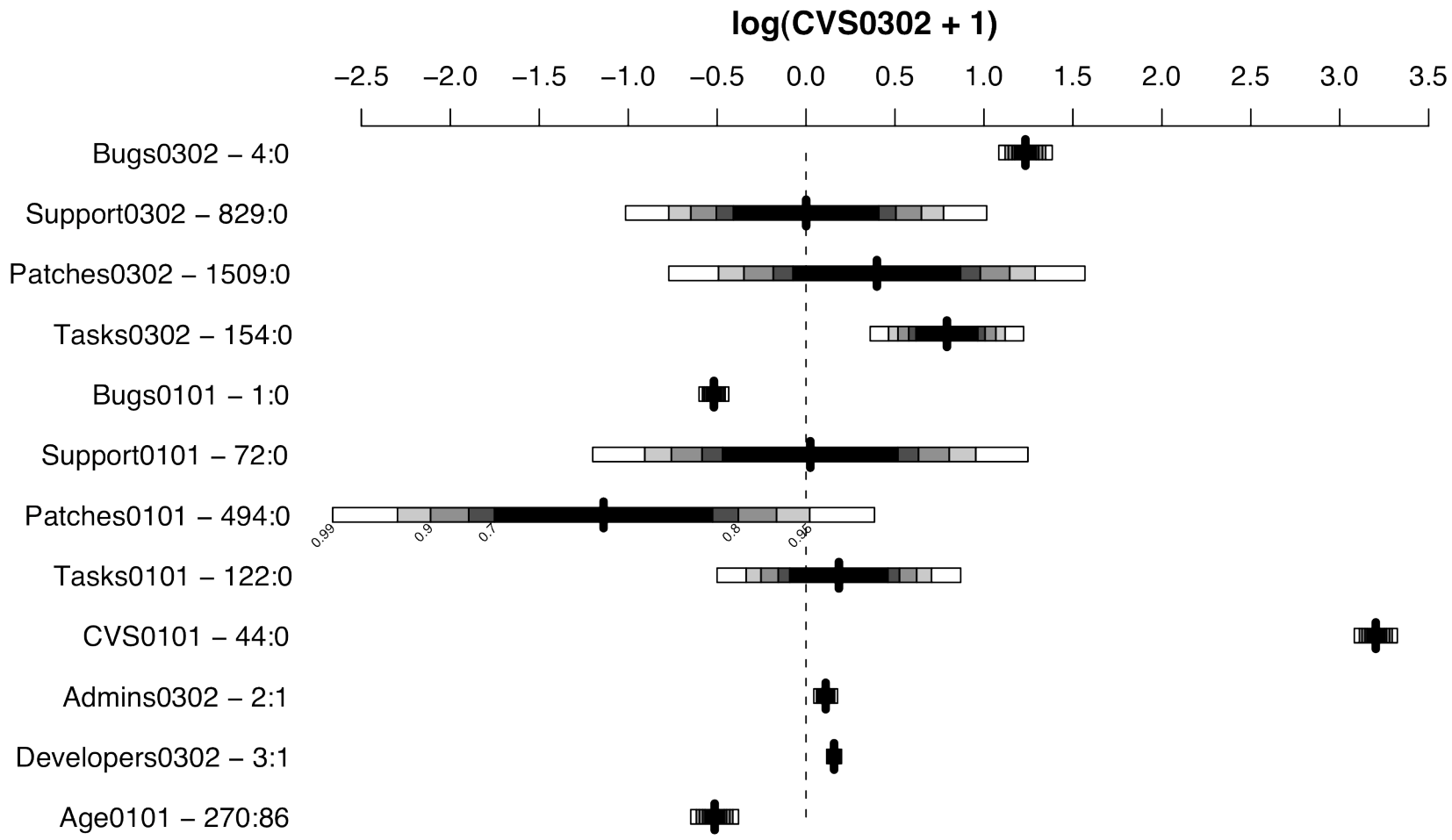
R^2=0.51

To sum up…



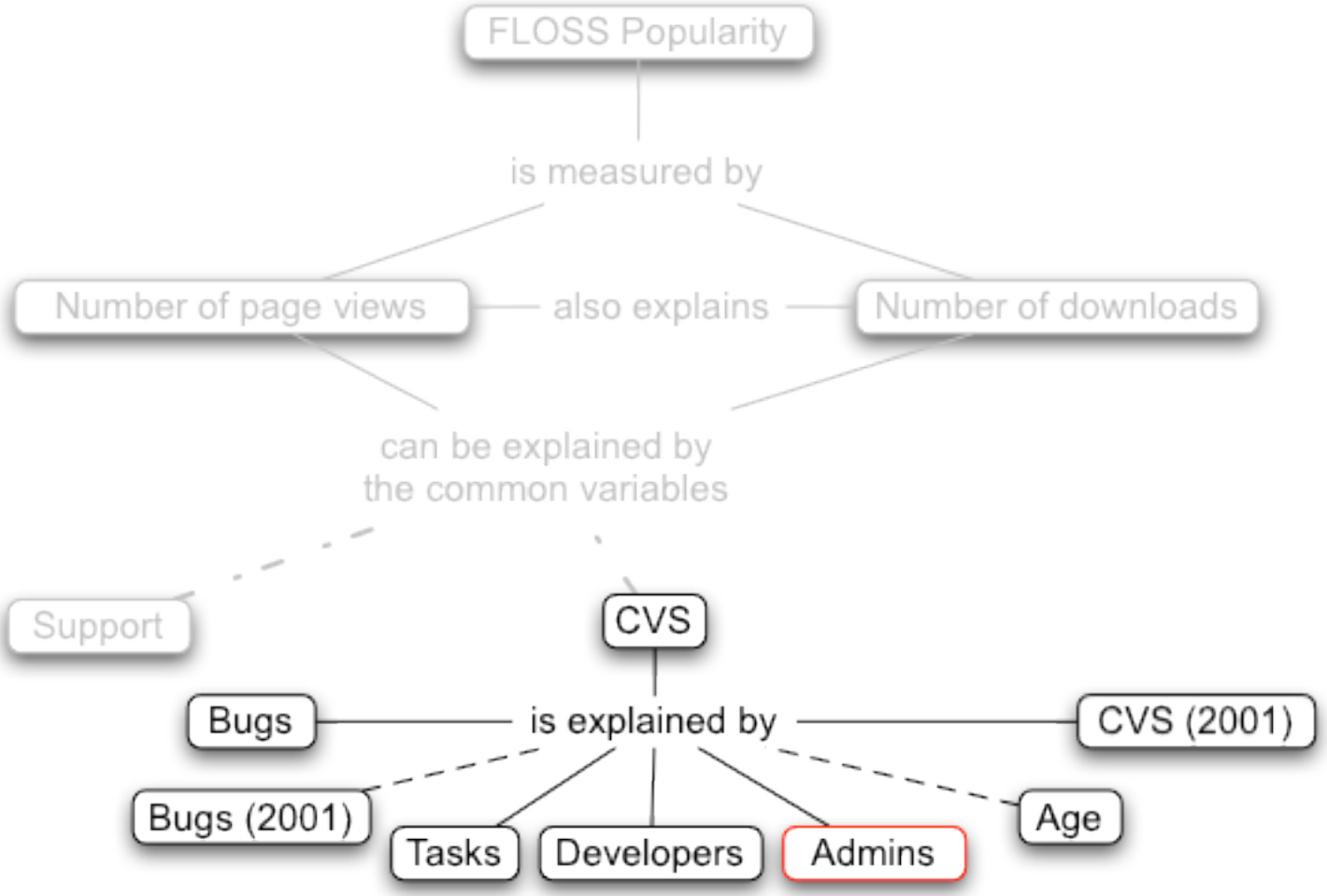
These variables are good candidates for use in data mining.

Regression Analysis Results (4)



R^2=0.67

To sum up…


FLOSS Popularity
is measured by
Number of page views
also explains
Number of downloads
can be explained by
the common variables
Support
CVS
Bugs
is explained by
CVS (2001)
Bugs (2001)
Tasks
Developers
Admins
Age

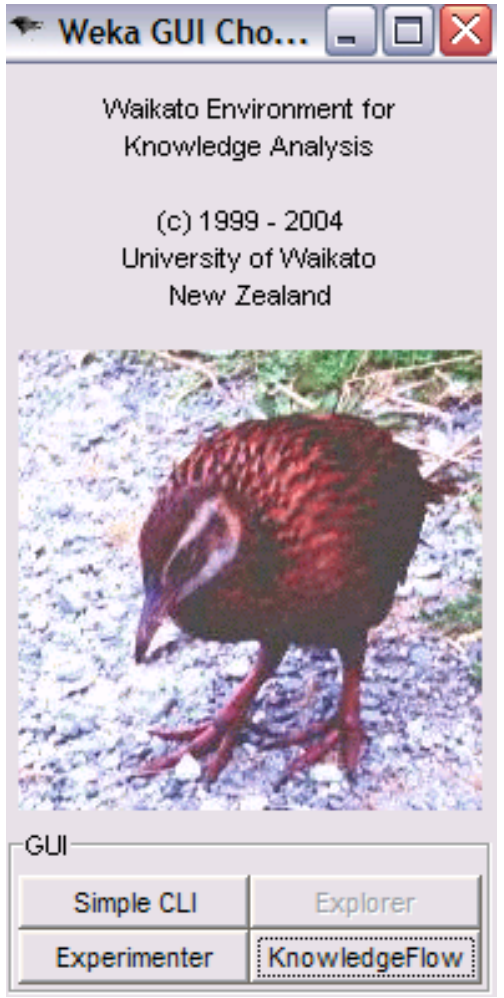
Data mining

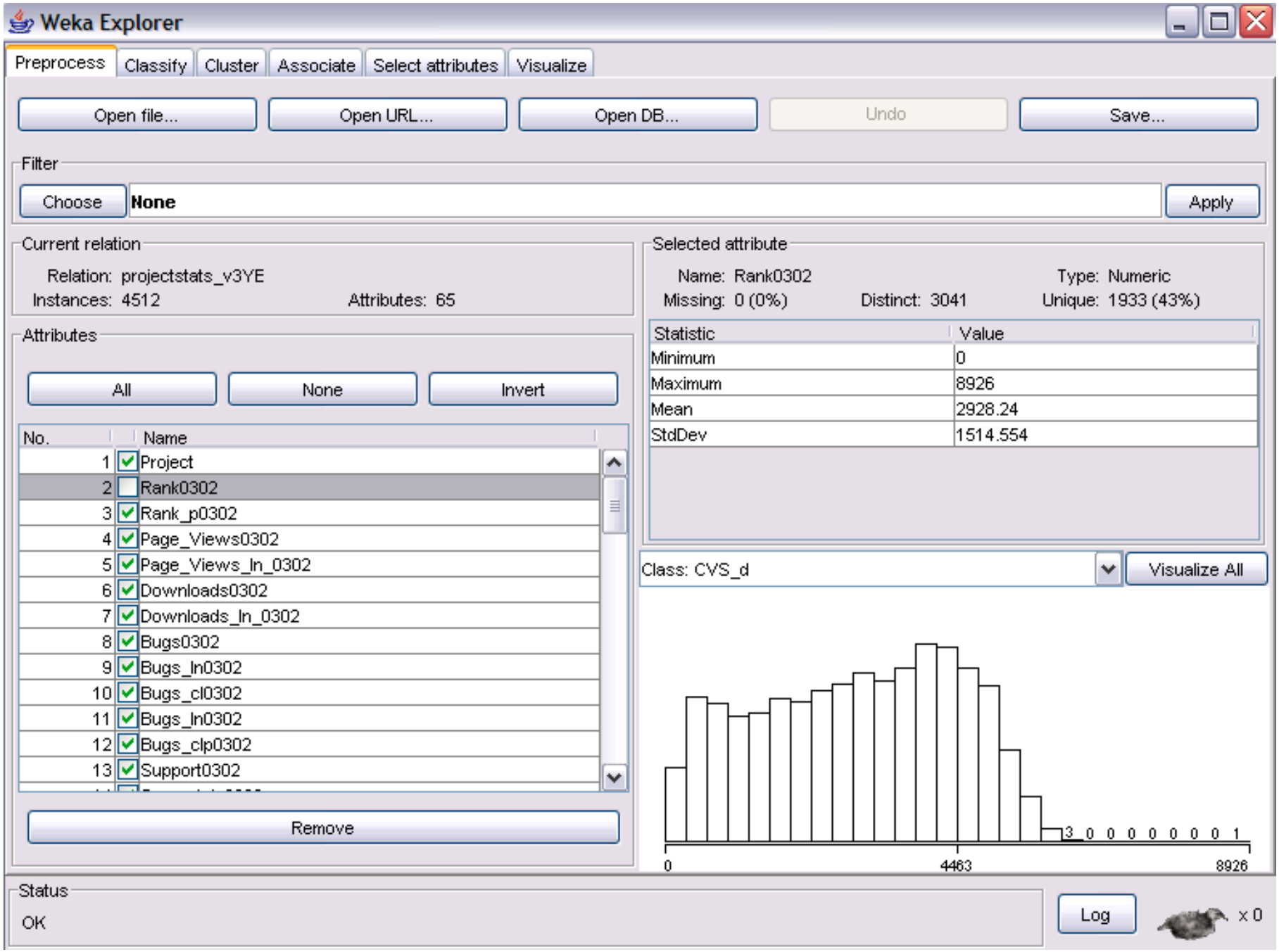
- Data mining algorithms can find useful and well-supported patterns in data that are otherwise hard to find*

*Source: Cooper, L. G., & Giuffrida, G. (2000). Turning datamining into a management science tool: New algorithms and empirical results. *Management Science, 46*(2), 249-264.

Data analysis tool

- Waikato Environment for Knowledge Analysis (Weka)
- Supports:
  - Processing and filtering of data
  - Visualization of the data using histograms and scatter plots
  - Analysis using classification, clustering and association rules
- Used Tertius method




Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: projectstats_v3YE
Instances: 4512
Attributes: 65
Selected attribute
Name: Rank0302
Type: Numeric
Missing: 0 (0%)
Distinct: 3041
Unique: 1933 (43%)
Statistic
Value
Minimum
0
Maximum
8926
Mean
2928.24
StdDev
1514.554
Attributes
All
None
Invert
No.
Name
1 Project
2 Rank0302
3 Rank_p0302
4 Page_Views0302
5 Page_Views_ln_0302
6 Downloads0302
7 Downloads_ln_0302
8 Bugs0302
9 Bugs_ln0302
10 Bugs_cl0302
11 Bugs_ln0302
12 Bugs_clp0302
13 Support0302
Remove
Class: CVS_d
Visualize All
3 0 0 0 0 0 0 0 1
0
4463
8926
Status
OK
Log
x 0

Data mining using association rules

- Association rules are used ideally in dealing with huge amounts of data, rather than as sensitive ways of testing and evaluating algorithms on limited datasets*
- Association method looks for rules that strongly associate different attribute values.
- Resulting rules are of the form:
  - If A and B are true (premise); then C can be expected (consequence)

*Source: Witten, I. H., & Frank, E. (1999). *Data mining: Practical machine learning tools and techniques with java implementations*.San Diego, CA: Academic Press.

Data limitations

- For association rule mining, data must be categorical
  - Rules are of the form “if A”, where “A” is that an attribute is in a particular category
  - Data had to be recoded into categories, usually low, medium, high

Statistics

- If A and B is true (support); then C can be expected with an n% accuracy (confidence)
- Support
  - # of cases for which the premise of rule holds
- Confidence
  - % of cases for which antecedent and consequent are both true
- True and false positive rates

Data Mining Results (1/7)

- If a project is inactive, then it's low ranked

Sup	TP	FP				
65%	92%	26%	Rank_p0302 = Lower_third	←	Patches_clp0101 = None and DLd = 1-1K
65%	92%	26%	Rank_p0302 = Lower_third	←	DLd = 1-1K
57%	82%	25%	Rank_p0302 = Lower_third	←	DL0302 = 1-1K
57%	82%	25%	Rank_p0302 = Lower_third	←	DL0302 = 1-1K and DL0101 = 1-1K

Data Mining Results (2/7)

- If a project's rank went up, typically it was young; if it was young, then generally its rank went up

Sup	TP	FP			
57%	87%	25%	Age0101 = 0-4_months	←	Rank_d = Up
58%	65%	9%	Rank_d = Up	←	Age0101 = 0-4_months

Data Mining Results (3/7)

- If a project isn't active, it has few downloads

Sup	TP	FP			
91%	99%	9%	DL0302 = 1-1K	←	DL0101 = 1-1K and DLd = 1-1K
89%	99%	10%	DL0302 = 1-1K	←	DLd = 1-1K
57%	72%	15%	DL0302 = 1-1K	←	Rank_p0302 = Lower_third
51%	51%	4%	DL0302 = 1-1K	←	PV0302 = 1-1K
50%	51%	4%	DL0302 = 1-1K	←	PVd = 1-1K
52%	83%	32%	DL0302 = 1-1K	←	Bugs_clp0302 = None and DL0101 = 1-1K
51%	76%	24%	DL0302 = 1-1K	←	Bugs0302 = None and DL0101 = 1-1K
51%	95%	46%	DL0302 = 1-1K	←	Support0302 = None and DL0101 = 1-1K
75%	85%	3%	DL0302 = 1K-10K	←	DLd = 1K-10K

Data Mining Results (4/7)

- If a project isn't active, it has few page views

Sup	TP	FP			
67%	86%	3%	PV0302 = 1-1K	←	PV0101 = 1-1K and PVd = 1-1K
51%	89%	26%	PV0302 = 1-1K	←	PV0101 = 1-1K and DLd = 1-1K
64%	85%	5%	PV0302 = 1-1K	←	PVd = 1-1K and DL0101 = 1-1K
64%	86%	6%	PV0302 = 1-1K	←	PVd = 1-1K
52%	88%	24%	PV0302 = 1-1K	←	DL0302 = 1-1K and PV0101 = 1-1K
50%	93%	32%	PV0302 = 1-1K	←	DL0302 = 1-1K and Patches_clp0302 = None
50%	93%	32%	PV0302 = 1-1K	←	DL0302 = 1-1K and Support_clp0302 = None
50%	90%	28%	PV0302 = 1-1K	←	DL0302 = 1-1K and Bugs_clp0101 = None
50%	93%	33%	PV0302 = 1-1K	←	DL0302 = 1-1K and Patches_clp0101 = None
50%	94%	33%	PV0302 = 1-1K	←	DL0302 = 1-1K and DL0101 = 1-1K
50%	93%	33%	PV0302 = 1-1K	←	DL0302 = 1-1K and Support_clp0101 = None
72%	80%	6%	PV0302 = 1K-10K	←	PVd = 1K-10K
72%	81%	0%	PV0302 = 10K+	←	PVd = 10K+
52%	90%	26%	PVd = 1-1K	←	PV0101 = 1-1K and DLd = 1-1K

Data Mining Results (5/7)

- If a project had a lot of CVS activity, then it still does, and vice versa

Sup	TP	FP			
71%	77%	0%	CVS0302 = 100+	←	CVSd = 100+
53%	53%	0%	CVS0302 = 100+	←	CVS0101 = 100+
97%	100%	0%	CVS0302 = None	←	CVS0101 = None and CVSd = None
77%	100%	21%	CVS0302 = None	←	CVSd = None
71%	100%	29%	CVS0302 = None	←	CVS0101 = None

Data Mining Results (6/7)

- Projects with one developer always have one administrator; and often vice versa

Sup	TP	FP		
61%	69%	0%	Admins0302 = 1.0	← Developers0302 = 1.0
59%	100%	43%	Developers0302 = 1.0	← Admins0302 = 1.0

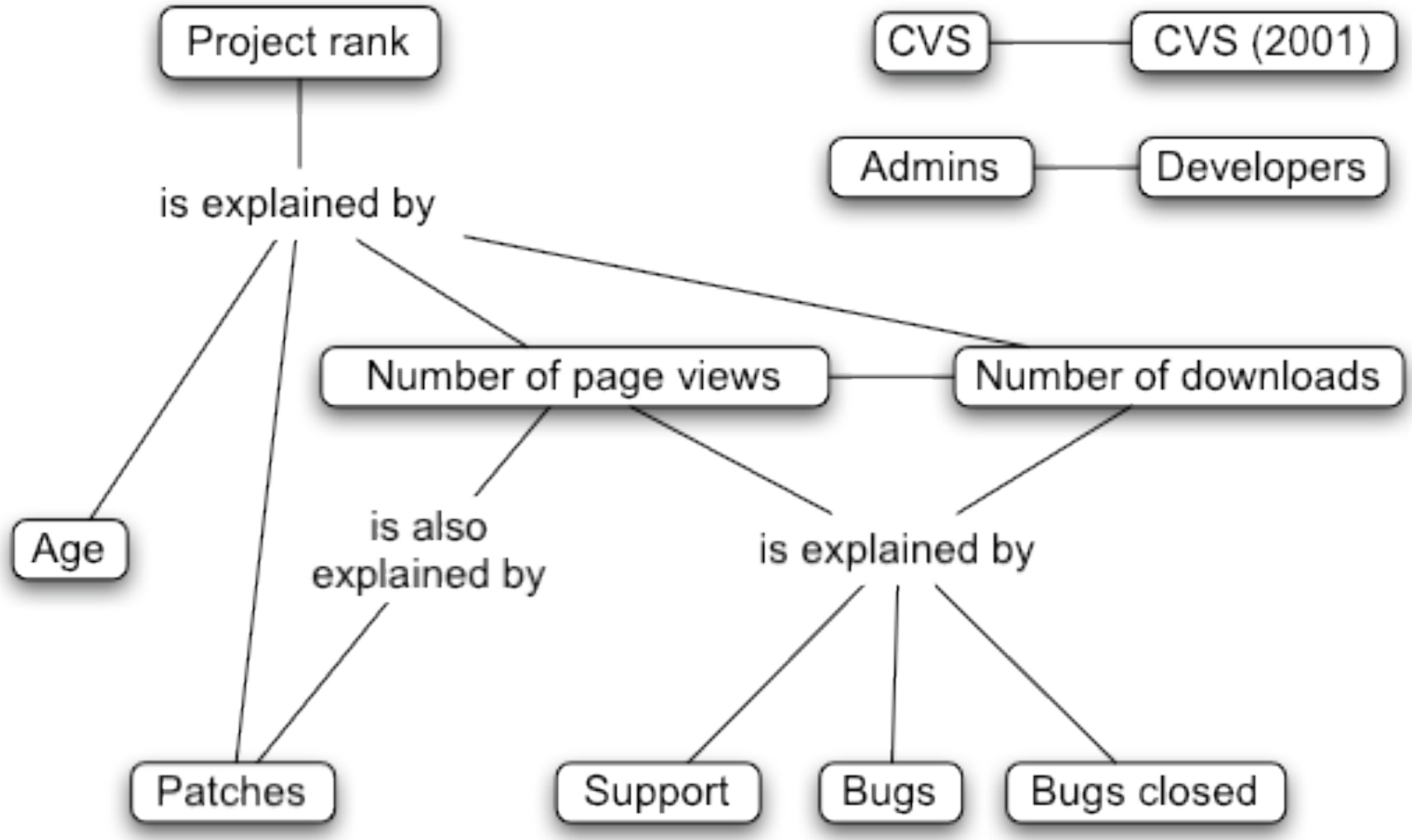
Data Mining Results (7/7)

- If a project doesn't use the SourceForge trackers, it doesn't

Sup	TP	FP			
79%	100%	20%	Bugs0302d = None	←	Bugs_clp0302d = None
68%	100%	33%	Bugs0302d = None	←	Bugs0101d = None
80%	85%	0%	Bugs_clp0302d = None	←	Bugs0302d = None
62%	100%	2%	Tasks0302d = None	←	Tasks_clp0302d = None
63%	100%	0%	Tasks_clp0302d = None	←	Tasks0302d = None

etc.

To sum up…


Project rank
CVS
CVS (2001)
Admins
Developers
is explained by
Number of page views
Number of downloads
Age
is also explained by
is explained by
Patches
Support
Bugs
Bugs closed

Summary

- Some clear relations among variables, but nothing too surprising
  - Most rules were obvious in retrospect
- “Missing mass”
  - Most rules were negative: they explained poor results
  - Topic does not help explain success
  - Number of developers does not help
    - did not appear in rules
    - low effect size in regression

Interesting projects to study

- Projects that showed a dramatic increase (Between 5 million and 8 million) in page views
  - Freenet, Phpbb, Alexandria, Blacknova
- Projects that showed a dramatic increase (Between 4 million and 8 million) in downloads
  - Cdexos, Virtualdub
- Projects that showed a dramatic increase (Between 8,200 and 9000) in rank
  - Tkvnc, Linux-ntfs, Thecob, Undb
- Projects that showed a dramatic decrease (Between 4,400 and 5,200) in rank
  - Lehtori, Gnap, Kpasman, Auit, Gnooutlook,Gwl

Future work

- Should employ better definitions of project effectiveness (not just downloads)
- Should use more data than just SourceForge demographics
- Should be based on theory

[686314] New incoming convo takes focus

You may monitor this Tracker item after you login (register an account, if you do not already have one).

Submitted By:
Joseph Trapasso (poppins24m)
Date Submitted:
2003-02-13 16:44
Last Updated By:
lschiere - Comment added
Date Last Updated:
2003-05-05 17:23
Category:
None
Group:
None
Assigned To:
Nobody/Anonymous (nobody)
Priority:
5
Status:
Open
Resolution:
Accepted

Summary:
New incoming convo takes focus

Alpha 5, Win2000
When I'm typing to a buddy (aim) and another buddy sends me the first message of a conversation, the focus is stolen from my text box and placed on the newly created tab. So, my continuing to type is futile.

I think this is new to Alpha 5.

Followups:

Message

Date: 2003-05-05 17:23
Sender: lschiere
Logged In: YES
user_id=28833

i think this might be fixed in 0.63

Date: 2003-04-05 16:18
Sender: hermanator
Logged In: YES
user_id=613964

Still is the case.. re-opening

Date: 2003-04-05 11:35
Sender: lschiere
Logged In: YES
user_id=28833

Gaim 0.60 has been released. This makes your bug report out of date.

Date: 2003-02-24 11:28
Sender: hermanator

Linux-Kernel Archive By Thread

Most Recent messages
1890 messages sorted by: [author] [date] [Subject]
About this archive
Other mail archives

- Re: [PATCH] API for true Random Number Generators to add entropy(2.6.11) *Jeff Garzik* (Thu Mar 24 2005 - 00:16:16 EST)
  - Re: [PATCH] API for true Random Number Generators to add entropy (2.6.11) *Folkert van Heusden* (Thu Mar 24 2005 - 07:38:55 EST)
  - Re: [PATCH] API for true Random Number Generators to add entropy (2.6.11) *David McCullough* (Thu Mar 24 2005 - 07:57:39 EST)
    - Re: [PATCH] API for true Random Number Generators to add entropy(2.6.11) *Jeff Garzik* (Thu Mar 24 2005 - 15:56:31 EST)
  - Re: [PATCH] API for true Random Number Generators to add entropy(2.6.11) *Jeff Garzik* (Thu Mar 24 2005 - 00:18:22 EST)
    - Re: [PATCH] API for true Random Number Generators to add entropy(2.6.11) *Andrew Morton* (Thu Mar 24 2005 - 00:36:27 EST)
      - Re: [PATCH] API for true Random Number Generators to add entropy (2.6.11) *Matt Mackall* (Mon Mar 28 2005 - 20:36:49 EST)
    - Re: [PATCH] API for true Random Number Generators to add entropy(2.6.11) *Randy.Dunlap* (Thu Mar 24 2005 - 00:45:39 EST)
  - Re: [PATCH] API for true Random Number Generators to add entropy(2.6.11) *Evgeniy Polyakov* (Thu Mar 24 2005 - 06:57:39 EST)


```
<?php
/**
 * abook_local_file.php
 *
 * Copyright (c) 1999-2005 The SquirrelMail Project Team
 * Licensed under the GNU GPL. For full terms see the file COPYING.
 *
 * @version $Id: abook_local_file.php,v 1.35 2005/03/20 10:06:45 tokul
$
 * @package squirrelmail
 * @subpackage addressbook
 */

/**
 * Backend for address book as a pipe separated file
* Stores the address book in a local file
 *
 * An array with the following elements must be passed to
 * the class constructor (elements marked ? are optional):
 *<pre>
 *   filename  => path to addressbook file
 * ? create    => if true: file is created if it does not exist.
 * ? umask     => umask set before opening file.
 * ? name      => name of address book.
*</pre>
* @package squirrelmail
 */
class abook_local_file extends addressbook_backend {
...
```

Hackman's Group Effectiveness Model