

End-User Programmers and their Communities: An Artifact-based Analysis

Kathryn T. Stolee, Sebastian Elbaum, and Anita Sarma
University of Nebraska–Lincoln

{kstolee, elbaum, asarma}@cse.unl.edu

September 22, 2011

This work is supported by the NSF GRFP under CFDA#47.076, NSF Award #0915526, and AFOSR Award #9550-10-1-0406.

Introduction

End User Programmers

People who engage in programming activities to support their hobbies and work.

Introduction

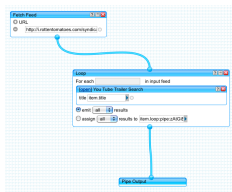
End User Programmers

People who engage in programming activities to support their hobbies and work.

	Professionals	End Users
Number in U.S.	3 million	13 million
Typical Education	C.S. Degree	Other Degree
Role of Programming	It's their job	It supports their job

Many Domains and Applications

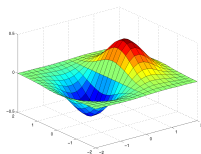
Web Mashups:



Educational Games:

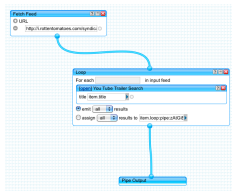


Scientific Computing:



Many Domains and Applications

Web Mashups:



Yahoo! Pipes

Artifacts

100,000

Participants

90,000

Educational Games:

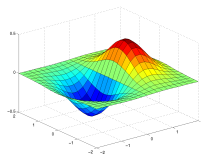


Scratch

700,000

500,000

Scientific Computing:



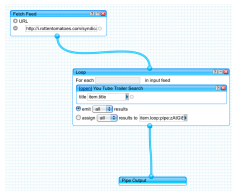
MATLAB

13,717

5,356

Many Domains and Applications

Web Mashups:



Yahoo! Pipes

Artifacts

100,000

Participants

90,000

Educational Games:

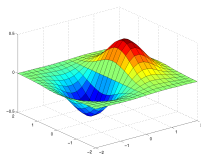


Scratch

700,000

500,000

Scientific Computing:



MATLAB

13,717

5,356

... yet we know little about **how** the repositories are utilized

Empirical Study Details

- **Research Goal**
- Study Context
- Research Questions
- Variables and Metrics
- Methods
- Results

Research Goal

To better understand *end-user programmer* communities

- Learn how communities and artifact repositories evolve
- Uncover needs for support in: development, maintenance, search, program understanding, . . .

Empirical Study Details

Goal: To better understand *end-user programmer* communities

- Research Goal
- **Study Context**
- Research Questions
- Variables and Metrics
- Methods
- Results

Why Mashup Communities?

Web Mashups

Applications that compose and manipulate existing data sources or services to create new data or service.

Why study mashups?

- Many environments (e.g., Apatar, DERI Pipes, IBM Mashup Center, Kivati, Yahoo! Pipes, . . .)
- Potential impact (many users, growth)

Why Mashup Communities?

Web Mashups

Applications that compose and manipulate existing data sources or services to create new data or service.

Why study mashups?

- Many environments (e.g., Apatar, DERI Pipes, IBM Mashup Center, Kivati, **Yahoo! Pipes**, ...)
- Potential impact (many users, growth)

About Yahoo! Pipes

The screenshot shows the Yahoo! Pipes web interface. At the top, there's a browser window with the address bar showing a URL to a pipe. Below the browser, the 'pipes' logo is visible along with navigation links: Home, My Pipes, Browse, Discuss, Documentation, and a 'Create a pipe' button. A search bar is also present. On the left sidebar, the user 'Kathryn' is logged in, with links to their profile, picture, and name. Below this are sections for 'Properties' (Not published, 0 clones), 'Bookmark / Share' (with social media icons), 'Tags (0)' (with an 'add new tag' button), 'Sources (2)' (listing 'google.com' and 'news.google.com'), and 'Modules (1)' (listing 'fetch', 'sort', and 'filter'). A diagram at the bottom left shows a sequence of three modules connected by arrows. The main content area displays the details for a pipe named 'example1', including its description, web address, and a list of fetched items. The first item is a news article about flooding in Pennsylvania, and the second is about a vote on replacing Anthony Weiner. A 'Report abusive Pipe' link is at the bottom right.

This example mashup
fetches and filters news
from `news.google.com`

Information page shows
the pipe output and
descriptive information

About Yahoo! Pipes

The screenshot shows the Yahoo! Pipes interface. At the top, there's a navigation bar with links: Home, My Pipes, Browse, Discuss, Documentation, and a 'Create a pipe' button. A search bar is also present. Below the navigation bar, the user profile for 'Kathryn' (Lincoln, Nebraska) is displayed. To the right, under the heading 'example1', there's a description of a pipe and a list of example pipes. The 'Publish' button is highlighted with a red box. Below the list, there's a 'Report abusive Pipe' link.

example1
Description of the pipe
Pipe Web Address: <http://pipes.yahoo.com/pipes/pipe.info?id=b587c0aa42daa5f67241b05fb4676a90> (edit)

☆ Edit Source Delete **Publish** Clone

Use this Pipe

Get as a Badge NY Yahoo! Google+ Get as RSS Get as JSON More options

List 3 items

Pennsylvania hit by huge flooding, towns submerged - Reuters
CTV.ca Pennsylvania hit by huge flooding, towns submerged Reuters WILKES-BARRE, Pa., Sept 9 (Reuters) - The remnants of Tropical Storm Lee eased their grip on Friday on Pennsylvania and other eastern US states after causing massive flooding but some towns remained submerged amid worry that river water ... As floodwaters recede tainted water a concern USA Today Counties in 'Triage Mode' as Flooding Persists in Northeast New York Times Days of rain trigger widespread Northeast floods The Associated...

Vote on replacing Anthony Weiner to test Democrats - Los Angeles Times
New York Daily News Vote on replacing Anthony Weiner to test Democrats Los Angeles Times New York's 9th Congressional District should be a seat that Democrats hold in the face of any disaster, man-made or natural -- but that political wisdom will be sorely tested next week when voters choose a ... Republicans look for upset in NY special election The Associated Press Israel and Obama: The Big Issues in the Race to Replace Anthony Weiner The Atlantic Poll shows tight race for Weiner's seat in...

Normal life starts resuming after power restored in San Diego - Los Angeles Times
Reuters Normal life starts resuming after power restored in San Diego Los Angeles Times Thirteen hours after an unprecedented outage left a huge swath of Southern California without power and brought normal life to a halt, power was restored to all 1.4 million customers of San Diego Gas & Electric, the biggest service area. ... San Diego Area Shakes Blackout TheStreet.com Power is Restored in San Diego Area New York Times Power back on for most in Arizona, California, Mexico Yuma Sun Bismarck Tribune...

[Report abusive Pipe](#)

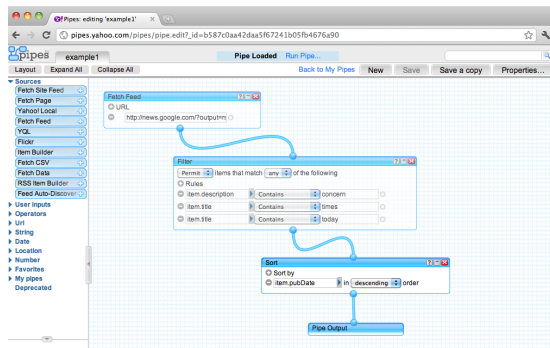
Clicking **Publish** adds the pipe to the public repository

About Yahoo! Pipes

The screenshot shows the Yahoo! Pipes interface. At the top, there's a navigation bar with links: Home, My Pipes, Browse, Discuss, Documentation, and a 'Create a pipe' button. A search bar is also present. Below the navigation bar, the user profile for 'Kathryn' (Lincoln, Nebraska) is displayed. To the right, a section titled 'example1' shows a list of pipes. The first pipe in the list is 'Pennsylvania hit by huge flooding, towns submerged - Reuters', which is highlighted with a red box around its 'Edit Source' button. Other pipes listed include 'Vote on replacing Anthony Weiner to test Democrats - Los Angeles Times' and 'Normal life starts resuming after power restored in San Diego - Los Angeles Times'.

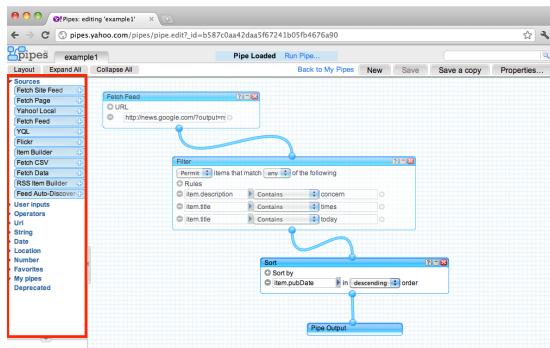
Clicking **Edit Source** loads
the Pipes Editor

About Yahoo! Pipes



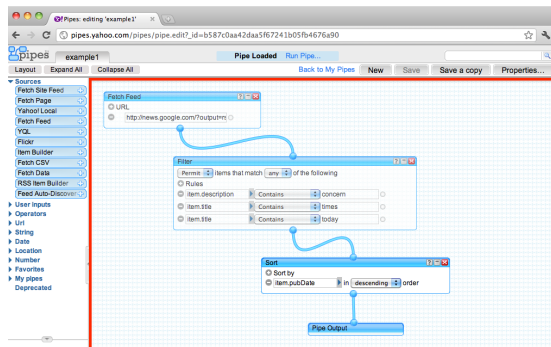
- Visual mashup creation environment
- Within a browser
- Drag and drop interface

About Yahoo! Pipes



- Visual mashup creation environment
- Within a browser
- Drag and drop interface

About Yahoo! Pipes



- Visual mashup creation environment
- Within a browser
- Drag and drop interface

About Yahoo! Pipes

The screenshot shows the Yahoo! Pipes web interface in a browser window. The address bar displays the URL: `pipes.yahoo.com/pipes/pipe.info?id=b587c0aa42daa5f67241b05fb4676a90`. The page header includes the 'pipes' logo and navigation links: Home, My Pipes, Browse, Discuss, Documentation, and a 'Create a pipe' button. A search bar is also present.

The main content area shows a pipe named 'example1' by user 'Kathryn' (Lincoln, Nebraska). The pipe's description is: 'Description of the pipe'. The pipe web address is: `http://pipes.yahoo.com/pipes/pipe.info?id=b587c0aa42daa5f67241b05fb4676a90`. Below the description are links to 'Edit Source', 'Delete', 'Publish', and 'Clone'.

The pipe's output is displayed as a list of 3 items. The first item is a news article titled 'Pennsylvania hit by huge flooding, towns submerged - Reuters'. The second item is a news article titled 'Vote on replacing Anthony Weiner to test Democrats - Los Angeles Times'. The third item is a news article titled 'Normal life starts resuming after power restored in San Diego - Los Angeles Times'.

On the left side of the interface, there are sections for 'Properties' (Not published, 0 clones), 'Bookmark / Share' (with social media icons), 'Tags' (add new tag), 'Sources' (google.com, news.google.com), and 'Modules' (fetch, sort, filter). At the bottom left, there is a diagram showing a sequence of modules connected by arrows.

- Visual mashup creation environment
- Within a browser
- Drag and drop interface

Empirical Study Details

Goal: To better understand *end-user programmer* communities

- Research Goal
- Study Context
- **Research Questions**
- Variables and Metrics
- Methods
- Results

Research Questions

RQ1: What are the characteristics of Yahoo! Pipes community?

1a,b: author attrition and author contributions

1c: artifact sharing, abstraction, complexity, and degree of overlap among pipes in the repository

Research Questions

RQ1: What are the characteristics of Yahoo! Pipes community?

1a,b: author attrition and author contributions

1c: artifact sharing, abstraction, complexity, and degree of overlap among pipes in the repository

RQ2: How do pipe attributes change as authors gain experience?

2a: experience measured by time

2b: experience measured by total contributions

Research Questions

RQ1: What are the characteristics of Yahoo! Pipes community?

1a,b: author attrition and author contributions

1c: artifact sharing, abstraction, complexity, and degree of overlap among pipes in the repository

RQ2: How do pipe attributes change as authors gain experience?

2a: experience measured by time

2b: experience measured by total contributions

RQ3: What are the characteristics of most prolific authors?

3a: author activities

3b: author skills

3c: awareness of the community

Research Questions

RQ1: What are the characteristics of Yahoo! Pipes community?

1a,b: author attrition and author contributions

1c: artifact sharing, abstraction, complexity, and degree of overlap among pipes in the repository

RQ2: How do pipe attributes change as authors gain experience?

2a: experience measured by time

2b: experience measured by total contributions

RQ3: What are the characteristics of most prolific authors?

3a: author activities

3b: author skills

3c: awareness of the community

Empirical Study Details

Goal: To better understand *end-user programmer* communities

- Research Goal
- Study Context
- Research Questions
- **Variables and Metrics**
- Methods
- Results

Study Details

Concept to Capture

artifact sharing/impact

abstraction

complexity

overlap of artifacts in repository

Variable

popularity

configurability

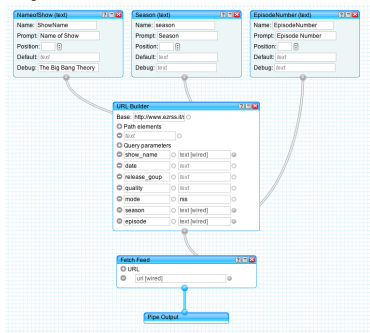
size

diversity

Study Details

Variables: size, configurability, popularity, diversity

Pipe Source



Pipe Information

Search EZTV for TV Shows

Searches EZTV's RSS for TV Shows that you specify and returns a RSS feed. Season and Episode Number not required.

Pipe Web Address: http://pipes.yahoo.com/pipes/pipe.info?_id=57d3c38b35160aee07cfabaf6d01b0229

☆ [View Source](#) [Clone](#)

Configure this Pipe

Episode Number

Season

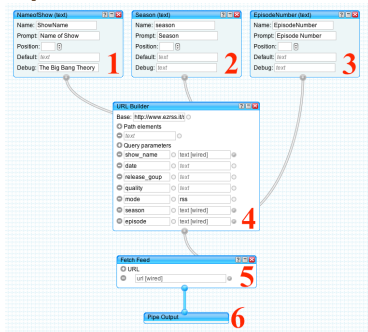
Name of Show

This Pipe may require all fields to have values before it will run successfully.
Please provide values into any empty field above and press "Run Pipe."

Study Details

Variables: size, configurability, popularity, diversity

Pipe Source



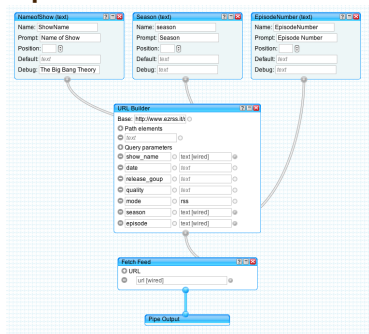
6 modules

Significance: Size is related to complexity

Study Details

Variables: size, **configurability**, popularity, diversity

Pipe Source



Pipe Information

Search EZTV for TV Shows

Searches EzTV's RSS for TV Shows that you specify and returns a RSS feed. Season and Episode Number not required.

Pipe Web Address: <http://pipes.yahoo.com/pipes/pipe.info?id=57d3c38b35160aee07cfabaf6d01b0229>

[View Source](#) [Clone](#)

Configure this Pipe

Episode Number

Season

Name of Show

This Pipe may require all fields to have values before it will run successfully.
Please provide values into any empty field above and press "Run Pipe."

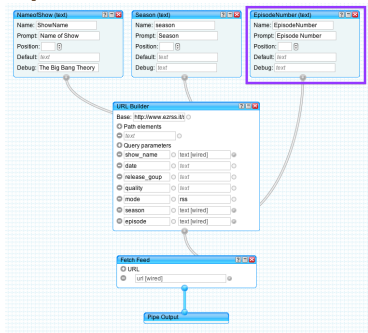
3 modules

Significance: Configurability is related to abstraction and language mastery

Study Details

Variables: size, **configurability**, popularity, diversity

Pipe Source



Pipe Information

Search EZTV for TV Shows

Searches EzTV's RSS for TV Shows that you specify and returns a RSS feed. Season and Episode Number not required.

Pipe Web Address: <http://pipes.yahoo.com/pipes/pipe.info?id=57d3c38b35160aee07cfabaf6d01b0229>

[View Source](#) [Clone](#)

Configure this Pipe

Episode Number

Season

Name of Show

This Pipe may require all fields to have values before it will run successfully.
Please provide values into any empty field above and press "Run Pipe."

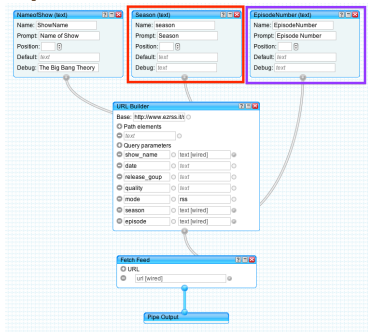
3 modules

Significance: Configurability is related to abstraction and language mastery

Study Details

Variables: size, **configurability**, popularity, diversity

Pipe Source



Pipe Information

Search EZTV for TV Shows

Searches EzTV's RSS for TV Shows that you specify and returns a RSS feed. Season and Episode Number not required.

Pipe Web Address: <http://pipes.yahoo.com/pipes/pipe.info?id=57d3c38b35160aee07cfabaf6d01b0229>

☆ [View Source](#) [Clone](#)

Configure this Pipe

Episode Number

Season

Name of Show [Run Pipe](#)

This Pipe may require all fields to have values before it will run successfully.
Please provide values into any empty field above and press "Run Pipe."

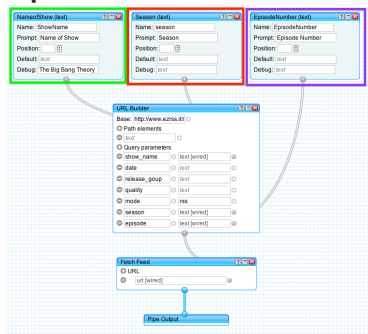
3 modules

Significance: Configurability is related to abstraction and language mastery

Study Details

Variables: size, **configurability**, popularity, diversity

Pipe Source



Pipe Information

Search EZTV for TV Shows

Searches EzTV's RSS for TV Shows that you specify and returns a RSS feed. Season and Episode Number not required.

Pipe Web Address: <http://pipes.yahoo.com/pipes/pipe.info?id=57d3c38b35160aee07cfabaf01b0229>

[View Source](#) [Clone](#)

Configure this Pipe

Episode Number

Season

Name of Show

This Pipe may require all fields to have values before it will run successfully.
Please provide values into any empty field above and press "Run Pipe."

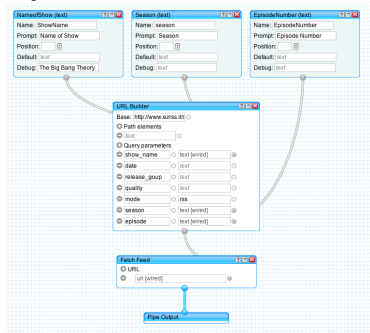
3 modules

Significance: Configurability is related to abstraction and language mastery

Study Details

Variables: size, configurability, **popularity**, diversity

Pipe Source



Pipe Information

Search EZTV for TV Shows

Searches EzTV's RSS for TV Shows that you specify and returns a RSS feed. Season and Episode Number not required.

Pipe Web Address: <http://ezraa.com/pipes/pipes/info?id=57d3c38b35160aee07cfabaf01b0229>


[View Source](#)
[Clone](#)

Configure this Pipe

 Episode Number

 Season

 Name of Show

This Pipe may require all fields to have values before it will run successfully.
Please provide values into any empty field above and press "Run Pipe."

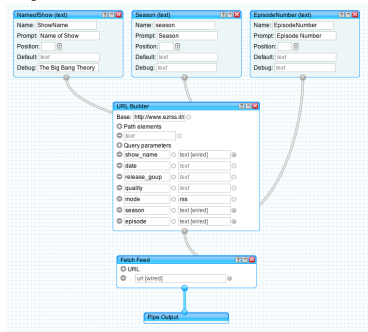
190 clones

Significance: Popularity is related to impact on community

Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



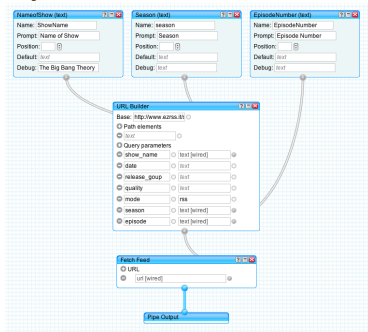
- 1 Same structure, fields, content
- 2 Same structure, field counts
- 3 Same structure
- 4 Same bag of modules
- 5 Same set of modules
- 6 Same type bag
- 7 Same size
- 8 No match

Significance: Diversity is related to contribution novelty

Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



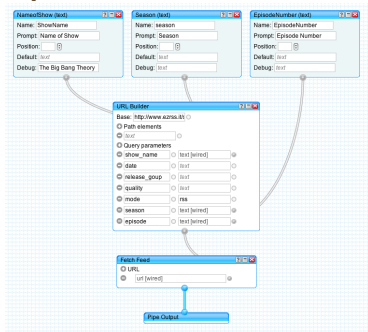
- 1 Same structure, fields, content
- 2 Same structure, field counts
- 3 **Same structure**
- 4 Same bag of modules
- 5 Same set of modules
- 6 Same type bag
- 7 Same size
- 8 No match

Significance: Diversity is related to contribution novelty

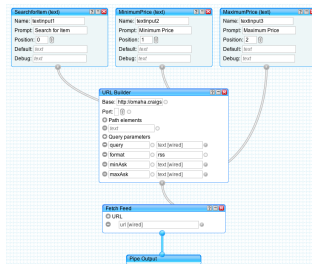
Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



3 Same structure

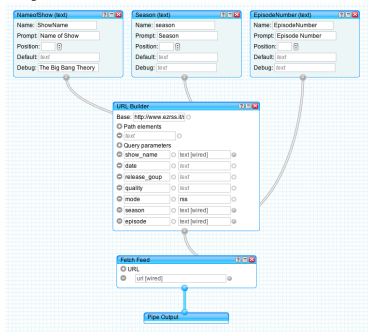


Significance: Diversity is related to contribution novelty

Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



- 1 Same structure, fields, content
- 2 Same structure, field counts
- 3 Same structure
- 4 Same bag of modules
- 5 Same set of modules
- 6 Same type bag
- 7 Same size
- 8 No match

Significance: Diversity is related to contribution novelty

Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



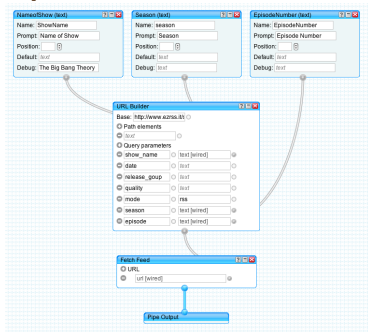
- 1 Same structure, fields, content
- 2 Same structure, field counts
- 3 Same structure
- 4 Same bag of modules
- 5 **Same set of modules**
- 6 Same type bag
- 7 Same size
- 8 No match

Significance: Diversity is related to contribution novelty

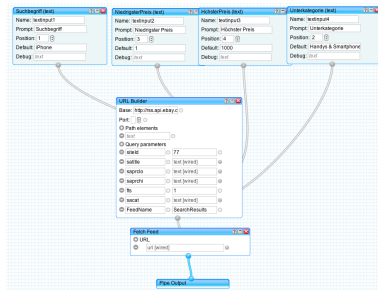
Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



5 Same set of modules

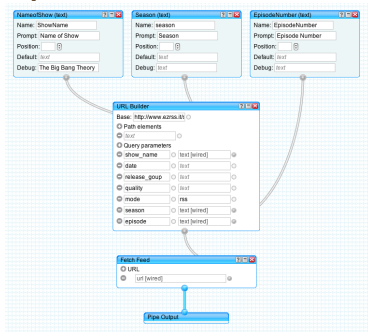


Significance: Diversity is related to contribution novelty

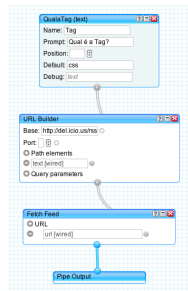
Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



5 Same set of modules

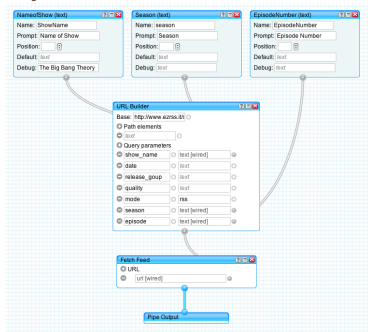


Significance: Diversity is related to contribution novelty

Study Details

Variables: size, configurability, popularity, **diversity**

Pipe Source



- 1 Same structure, fields, content
- 2 Same structure, field counts
- 3 Same structure
- 4 Same bag of modules
- 5 Same set of modules
- 6 Same type bag
- 7 Same size
- 8 No match

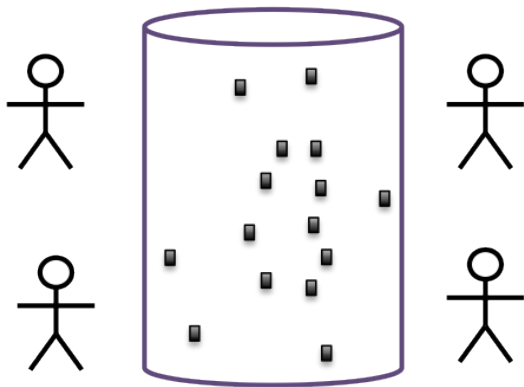
Significance: Diversity is related to contribution novelty

Empirical Study Details

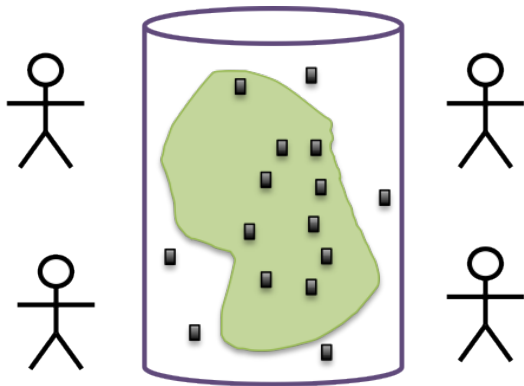
Goal: To better understand *end-user programmer* communities

- Research Goal
- Study Context
- Research Questions
- Variables and Metrics
- **Methods**
- Results

Data Collection

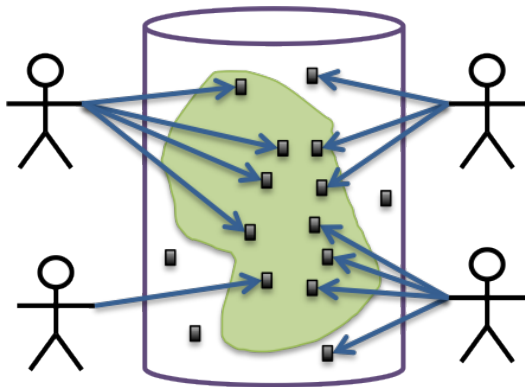


Data Collection



Artifacts: 32,887

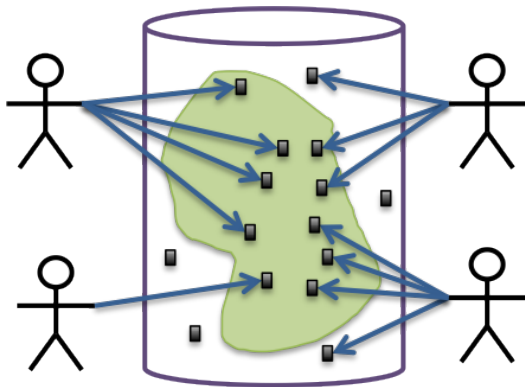
Data Collection



Artifacts: 32,887

Authors: 20,313

Data Collection



Artifacts: 32,887

Authors: 20,313

Threats: public repository offers limited visibility (internal); sampling bias (external); generalizability to other domains (external)

Empirical Study Details

Goal: To better understand *end-user programmer* communities

- Research Goal
- Study Context
- Research Questions
- Variables and Metrics
- Methods
- **Results**

Research Questions

RQ1: What are the characteristics of Yahoo! Pipes community?

1a,b: author attrition and author contributions

1c: artifact sharing, abstraction, complexity, and degree of overlap among pipes in the repository

RQ1: Characteristics of Yahoo! Pipes Community

Summary

Metric	Average
Size	8.20 modules per pipe
Configurability	0.65 modules per pipe
Popularity	5.67 clones per pipe
Diversity	3.62 cluster level

RQ1: Characteristics of Yahoo! Pipes Community

Summary

Metric	Average
Size	8.20 modules per pipe
Configurability	0.65 modules per pipe
Popularity	5.67 clones per pipe
Diversity	3.62 cluster level

34% of pipes are configurable

RQ1: Characteristics of Yahoo! Pipes Community

Summary

Metric	Average
Size	8.20 modules per pipe
Configurability	0.65 modules per pipe
Popularity	5.67 clones per pipe
Diversity	3.62 cluster level

54% of pipes have been cloned

RQ1: Characteristics of Yahoo! Pipes Community

Summary

Metric	Average
Size	8.20 modules per pipe
Configurability	0.65 modules per pipe
Popularity	5.67 clones per pipe
Diversity	3.62 cluster level

5% of pipes are exact duplicates, yet 46% have a match if field values are relaxed

RQ1: Characteristics of Yahoo! Pipes Community

Take Aways:

- There is a lot of reuse of shared pipes
- Participants often submit pipes that are highly similar to other pipes in the repository

Research Questions

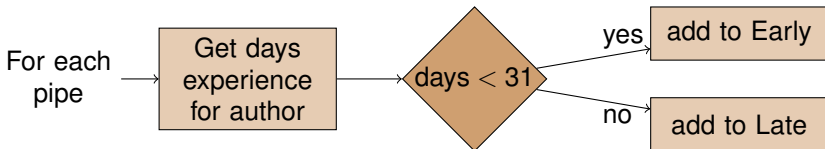
RQ2: How do pipe attributes change as authors gain experience?

2a: measures experience in terms of time

2b: measures experience in terms of total contributions

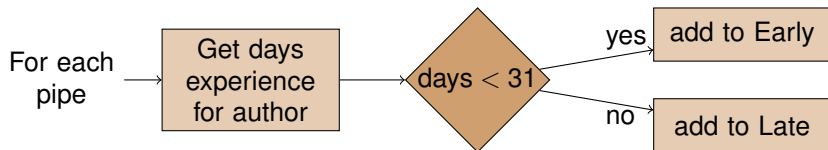
RQ2: Analysis of artifacts as authors gain experience

Comparisons based on experience (time)



RQ2: Analysis of artifacts as authors gain experience

Comparisons based on experience (time)



Characteristic	μ_{early}	μ_{late}
# of Pipes	27,555	5,332
Diversity***	3.519	4.126
Popularity***	4.984	9.254
Configurability***	0.614	0.838
Size***	7.919	9.587

$$H_0 : \mu_{early} > \mu_{late}$$

$$H_a : \mu_{early} \leq \mu_{late}$$

Signif. codes:

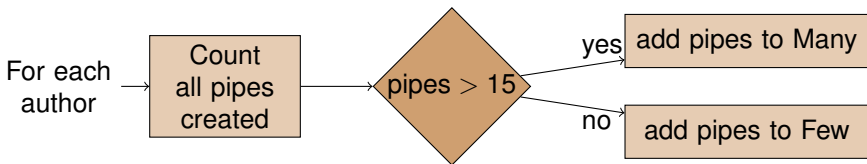
*** 0.001 ** 0.01

RQ2: Analysis of artifacts as authors gain experience

Take Away: More experience results in pipes that are larger, more popular, more configurable, and more diverse

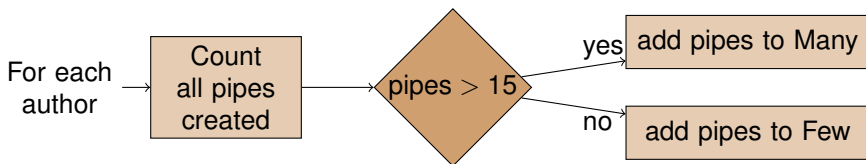
RQ2: Analysis of artifacts as authors gain experience

Comparisons based on contributions



RQ2: Analysis of artifacts as authors gain experience

Comparisons based on contributions



Characteristic	μ_{few}	μ_{many}
# of Pipes	30,503	2,384
Diversity	3.639	3.355
Popularity***	4.302	23.250
Configurability***	0.644	0.729
Size**	8.194	8.136

$$H_0 : \mu_{few} > \mu_{many}$$

$$H_a : \mu_{few} \leq \mu_{many}$$

Signif. codes:

*** 0.001 ** 0.01

RQ2: Analysis of artifacts as authors gain experience

Take Away: The most prolific authors create pipes that are larger, more popular, and more configurable

... what about diversity?

Research Questions

RQ3: What are the characteristics of most prolific authors?

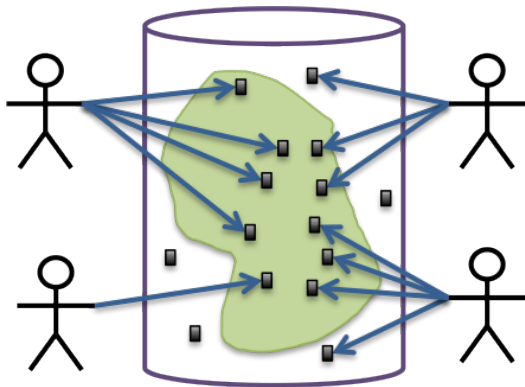
3a: author activities

3b: author skills

3c: awareness of the community

RQ3: Characteristics of most prolific authors

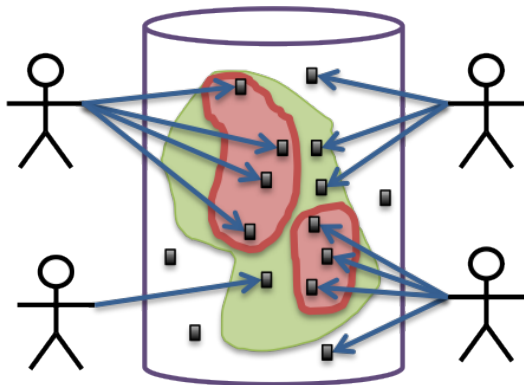
Study Set-up



Authors: 20,313

RQ3: Characteristics of most prolific authors

Study Set-up

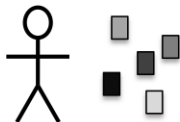


Authors: 20,313

Prolific Authors: 81

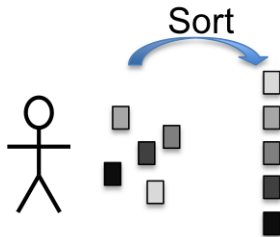
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



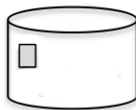
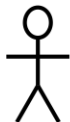
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



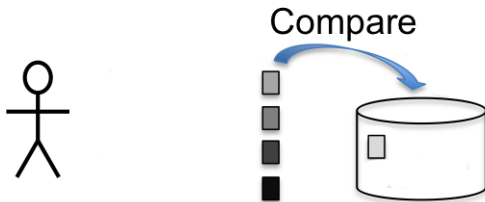
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



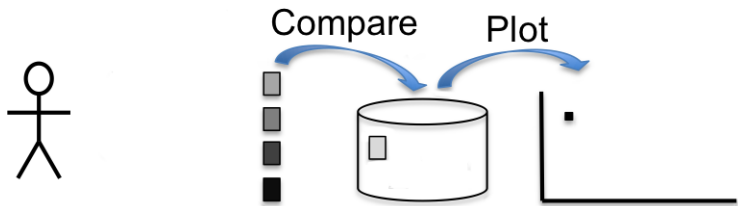
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



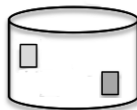
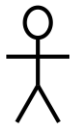
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



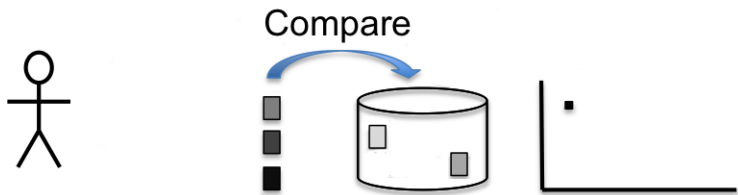
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



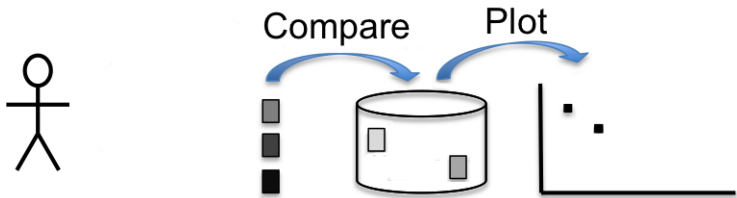
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



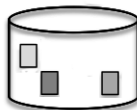
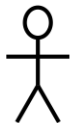
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



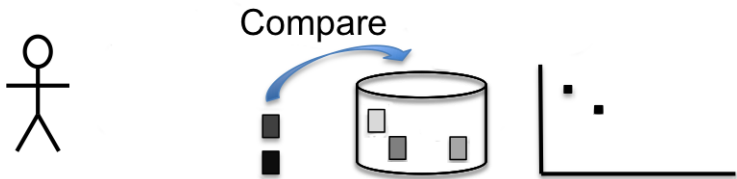
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



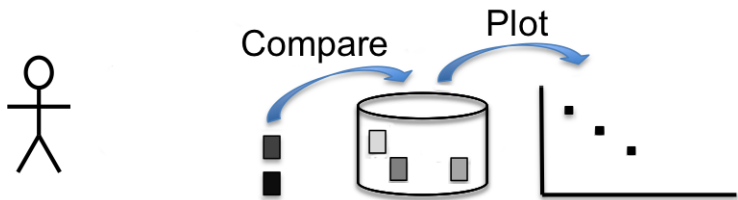
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



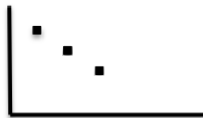
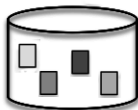
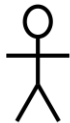
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



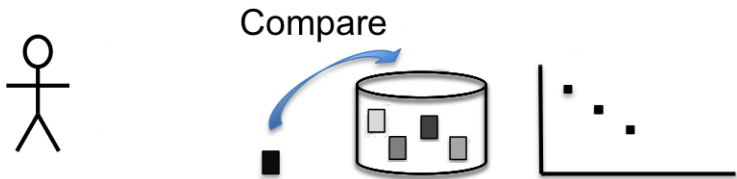
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



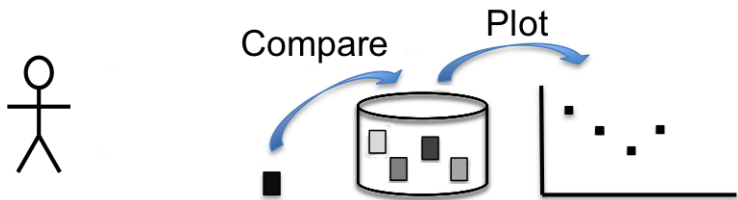
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



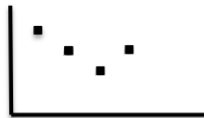
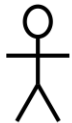
RQ3: Characteristics of most prolific authors

Rolling Cluster Analysis



RQ3: Characteristics of most prolific authors

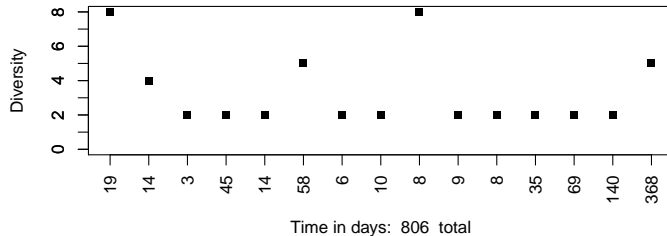
Rolling Cluster Analysis



RQ3: Characteristics of most prolific authors

Author Activities

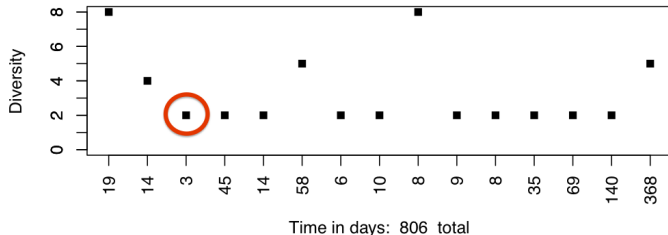
Rolling Diversity Analysis Over Time



RQ3: Characteristics of most prolific authors

Author Activities

Rolling Diversity Analysis Over Time



Level 2:
Same
structure and
field counts;
relax field
values

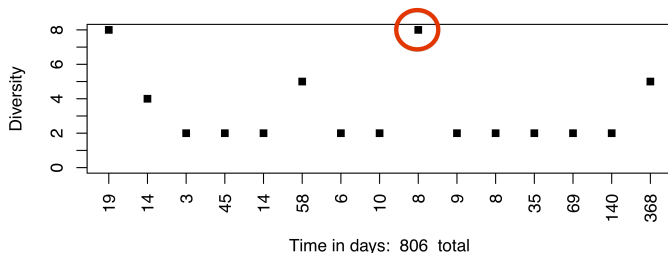
43% of pipes submitted by prolific authors represent **tweaks**

For Example: Change a URL, filter criterion, sort order, ...

RQ3: Characteristics of most prolific authors

Author Activities

Rolling Diversity Analysis Over Time



Level 8: No structural similarities

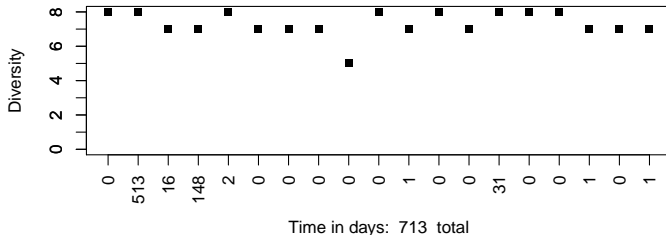
43% of pipes submitted by prolific authors represent **tweaks**

52% of pipes submitted by prolific authors represent **new initiatives**

RQ3: Characteristics of most prolific authors

Author Activities

Rolling Diversity Analysis Over Time

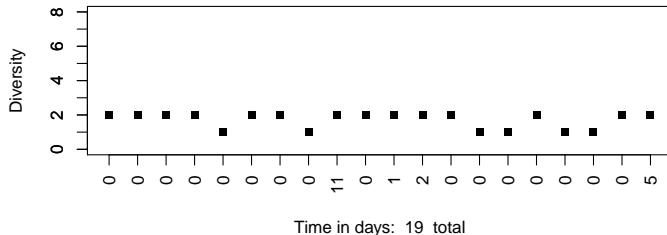


56% of prolific authors consistently submit **new initiatives**

RQ3: Characteristics of most prolific authors

Author Activities

Rolling Diversity Analysis Over Time



27% of prolific authors consistently submit **tweaks**

RQ3: Characteristics of most prolific authors

Take Away #1: 1/2 of participants submit pipes that are novel to their previous contributions

Take Away #2: 1/4 of participants submit pipes that are tweaks of their other pipes

The *real* take away

End-user programmer communities may need ...

- moderators.

- Repository is cluttered with highly similar artifacts (RQ1)

The *real* take away

End-user programmer communities may need ...

- moderators.
 - Repository is cluttered with highly similar artifacts (RQ1)
- more sophisticated repository search.
 - Many pipes are very structurally similar to other pipes in the repository (RQ1)
 - Early authors create less diverse pipes than later authors (RQ2)

The *real* take away

End-user programmer communities may need . . .

- moderators.
 - Repository is cluttered with highly similar artifacts (RQ1)
- more sophisticated repository search.
 - Many pipes are very structurally similar to other pipes in the repository (RQ1)
 - Early authors create less diverse pipes than later authors (RQ2)
- artifact development support.
 - Tweaks represent missed opportunities for parameterization (RQ3)
 - Many shared pipes are tweaks on previously-committed pipes by the same author (RQ3)

Threats to Validity

■ Internal

- History (the pipes were sampled at different times)
- Selection (the repository only provides public pipes)

■ Construct

- Interaction of different factors
- Mono-method bias on diversity (only consider structural diversity, not semantic)

■ External

- Generalizability (only studied one community)
- Sampling bias (could not control search results when sampling)

Conclusion

- Authors utilize the repository in different ways
- As authors gain experience in the environment, they tend to make more valuable contributions to the repository
- There is a need for better support to help end-user programmer communities continue to progress and grow
- To generalize the results, we are interested in extending the metrics to other languages and repositories

To facilitate replication, the data used in this analysis is available:

<http://cse.unl.edu/~kstolee/esem2011/artifacts.html>

Backup

Future Work

Extend this study to other end-user programming communities.

Metric considerations:

Size: requires some measure of program size and/or complexity

Configurability: requires some notion of abstraction or measure of configuration

Popularity: requires some notion of clones, copies, favorites, etc. to indicate community approval

Diversity: For structured languages, this metric can be adapted by mapping fields and modules to the language-specific structures

RQ1: Characteristics of Yahoo! Pipes Community

RQ1a: Attrition of contributors

82% of the 20,313 authors were active for only one day

RQ1b: Author contributions

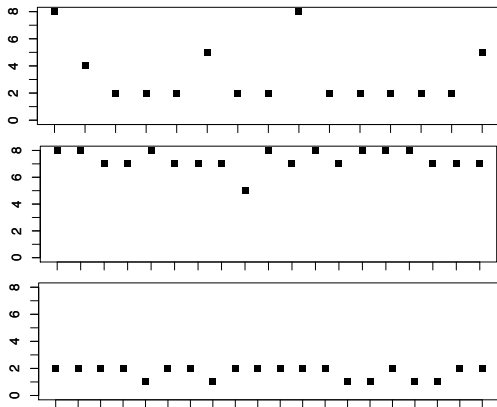
76% of authors submitted only one pipe

24% of authors submitted 53% of the pipes

Take Away: This end-user community has attrition and contribution patterns similar to other online communities

RQ3: Characteristics of most prolific authors

RQ3b: Author Skills



17.2% of prolific authors exhibit a **variable** skills pattern

55.6% of prolific authors exhibit a **high** skills pattern

27.2% of prolific authors exhibit a **low** skills pattern

Take Away: Authors utilize the repository in different ways

RQ3: Characteristics of most prolific authors

RQ3c: Author Awareness

Two types of awareness:

Local : awareness of author's previous contributions

Community : awareness of pipes in community

Local		Community	Awareness	Avg. % of Pipes
low	=	low	None	50.06%
*	>	*	Local	19.20%
high	=	high	Community	30.74%

Take Away: Authors have varying levels of community awareness