

Make it with friends

Break it yourself

Open-source software and DH research

Andrew J. Roback

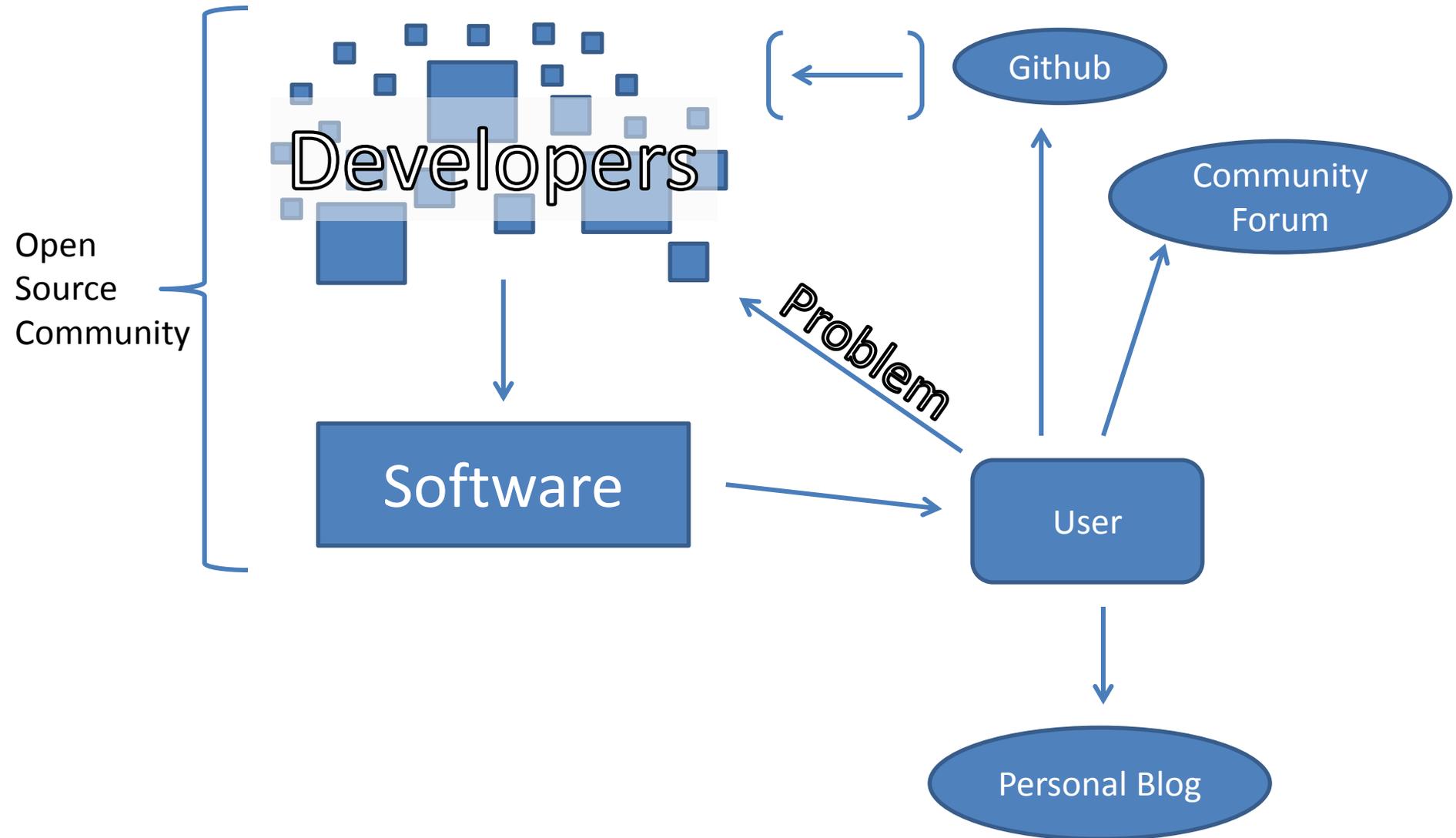
PhD Candidate, IIT

Open-source software

1. Free redistribution
2. Source code open to user
3. License allows for derived works
- ...
6. No discrimination against field or endeavor

[The Open Source Definition](#)

A path for open source work



Breaking it

Two definitions of “break”:

:(

1 e : to render inoperable

!!!

15 a : to find an explanation or solution for : SOLVE

Some examples

1. Transforming JSON to CSV
2. Automating web search using Watir
3. Creating a complex survey using LimeSurvey

JSON to CSV

Problem:

I have a lot of data from Twitter in JSON, but to load it into my survey instrument and analyze it, I need it in CSV format.

Brute force:

Manually edit hundreds of thousands of lines of code. Finish project next century.

JSON to CSV

JSON

- JavaScript Object Notation
- Open standard, human readable
- Sequence of nested, attribute-value pairs
- [Example](#)

JSON to CSV

CSV

- Comma separated values file
- Uses delimiters and line termination characters
- Easily imported into analysis software (e.g. Excel)

```
1 19563014,358241939411304449,"What do those food labels really mean? Check  
out @NRDCFood new app to help you navigate which claims are real.  
http://t.co/obcC514uIF",Fri Jul 19 15:08:28 +0000 2013,1,0,<a  
href=http://www.hootsuite.com rel=nofollow>HootSuite</a>  
2 19563014,355422478471925760,"Marco Foster will race on a Big Wheel for One  
Hope United as part of the Tour de Grid fundraiser for local...  
http://t.co/170v5rQdIX",Thu Jul 11 20:24:56 +0000 2013,0,0,<a  
href=http://www.facebook.com/twitter rel=nofollow>Facebook</a>
```

JSON to CSV

Solution

- Ruby (open-source programming language)
- JSON gem (a library of methods for transforming JSON)
- [Script](#)

JSON to CSV

Broke it :(

- When I didn't anticipate invalid JSON in my sample

Broke it!!!

- Discovered [JSON Lint Validator](#) and corrected with text editor
- Shared it: [repo on GitHub](#)

Automating web search

Problem:

I need to find social networking site (SNS) accounts for a list of people/organizations, but no such list exists. Also, Google search API deprecated.

Brute force:

Manually search several thousand terms and copy and paste URLs of accounts.

Automating web search

Solution

- Again, Ruby
- Watir gem
 - Made for website testing.
 - Automates interaction with web forms (lets you enter text, push buttons, etc.).
 - Works really well with arrays and loops
- Script and example

Automating web search

Broke it :(

- When I didn't anticipate Google changing the HTML structure of their search/results pages

Broke it!!!

- When I used XPath (query language for elements in a structured document) to locate elements I wanted browser to find and manipulate.
- Shared it: [personal blog](#)

Creating a complex survey

Problem:

I want to show survey participants their own SNS posts in a survey and have them describe their motivations for posting. Free platform doesn't exist.

Brute force:

Pay \$\$\$\$\$\$ for a platform that can do this.

Creating a complex survey

Solution

- LimeSurvey
 - Open-source survey platform built on PHP (same scripting language as WikkaWiki, WordPress, etc.)
- Use tokens (list of users) and associated user-defined fields to add customized information

Creating a complex survey

Goal

- Show users their own SNS data
 - Manipulation of files controlling tokens

Creating a complex survey

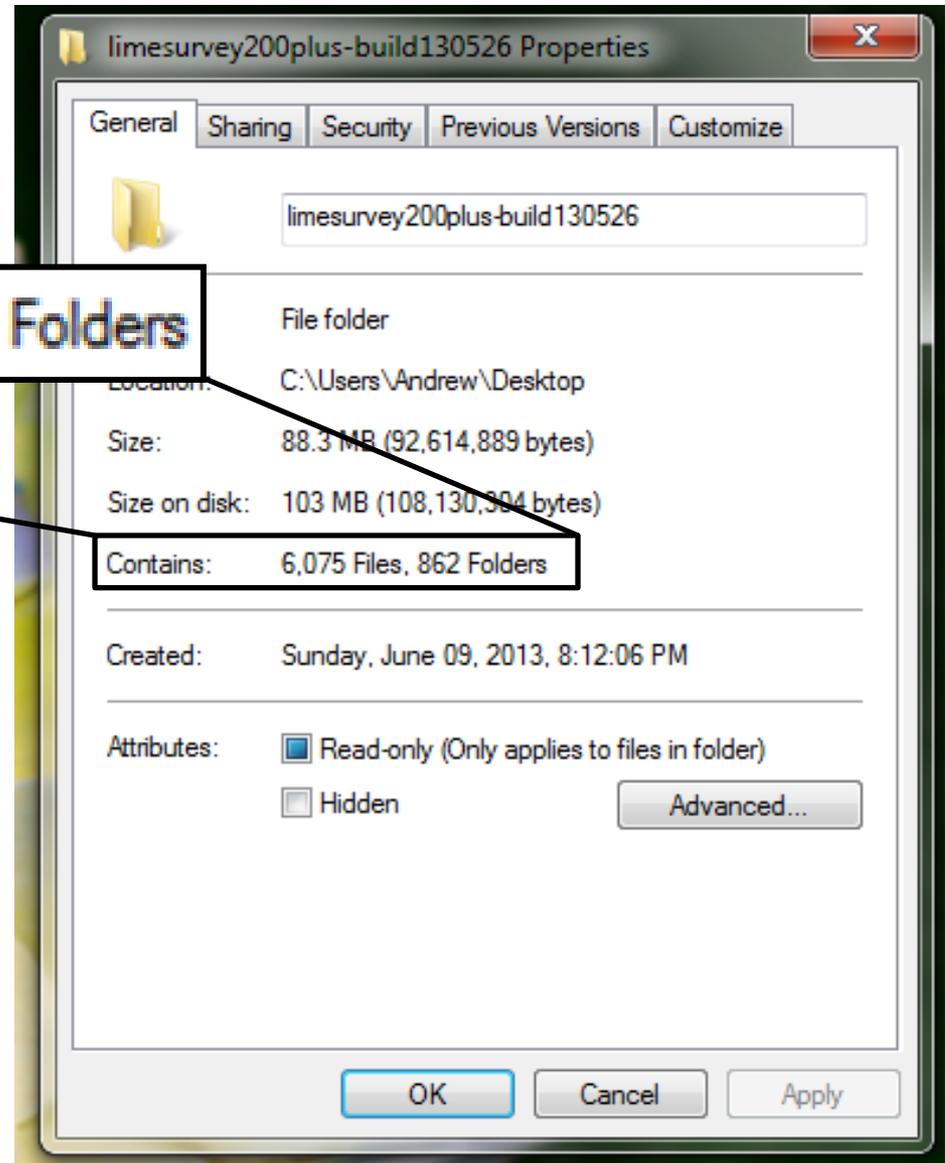
Broke it :(

- Facebook posts were getting truncated so users would not see the whole post.
- After some analysis: all truncated posts are 255 characters long
- Speculation: software is limiting user-defined token field imports by character length

Creating a complex survey

A challenge:

Contains: 6,075 Files, 862 Folders



Creating a complex survey

Free-source to the rescue

- Notepad ++
 - Lightweight, but very powerful text editor

Creating a complex survey

Broke it!!!

```
1093 // CHECK TO SEE IF A TOKEN TABLE EXISTS FOR THIS SURVEY
1094 $bTokenExists = tableExists('{{tokens_' . $iSurveyId . '}}');
1095 if (!$bTokenExists) //If no tokens table exists
1096 {
1097     self::_newtokentable($iSurveyId);
1098 }
1099
1100 $number2add = sanitize_int(Yii::app()->request->getPost('addnumber'), 1, 100);
1101 $tokenattributefieldnames = getAttributeFieldNames($iSurveyId);
1102 $i = 1;
1103
1104 for ($b = 0; $b < $number2add; $b++)
1105 {
1106     while (in_array('attribute_' . $i, $tokenattributefieldnames) !== false)
1107     {
1108         $i++;
1109     }
1110     $tokenattributefieldnames[] = 'attribute_' . $i;
1111     Yii::app()->db->createCommand(Yii::app()->db->getSchema()->addColumn('{{tokens_' . intval($iSurveyId) . '}}', 'attribute_' . $i, 'VARCHAR(500)')->execute();
1112     $fields['attribute_' . $i] = array('type' => 'VARCHAR', 'constraint' => '500');
1113 }
1114
1115 LimeExpressionManager::SetDirtyFlag(); // so that knows that token tables have changed
1116
1117 Yii::app()->session['flashmessage'] = sprintf($clang->gT("%s field(s) were successfully added."), $number2add);
1118 Yii::app()->getController()->redirect(Yii::app()->getController()->createUrl("/admin/tokens/sa/managetokenattributes/surveyid/$iSurveyId"));
1119
1120 }
1121
```

```
Find result - 4459 hits
Search "255" (4459 hits in 244 files)
C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\export.php (4 hits)
C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\remotecontrol.php (3 hits)
C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\tokens.php (2 hits)
Line 1115:     Yii::app()->db->createCommand(Yii::app()->db->getSchema()->addColumn('{{tokens_' . intval($iSurveyId) . '}}', 'attribute_' . $i, 'VARCHAR(255)')->execute();
Line 1116:     $fields['attribute_' . $i] = array('type' => 'VARCHAR', 'constraint' => '255');
C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\VerificationController.php (5 hits)
C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\statistics\helpers.php (8 hits)
```

Creating a complex survey

Broke it!!!

```
1093 // CHECK TO SEE IF A TOKEN TABLE EXISTS FOR THIS SURVEY
1094 $bTokenExists = tableExists('{{tokens_'. $iSurveyId .'}}');
C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\tokens.php
Line 1115:()->addColumn('{{tokens_'.intval($iSurveyId).'}}', 'attribute_' . $i, 'VARCHAR(255)')->execute();
Line 1116: 'constraint' => '255');

1100 $number2add = sanitize_int(Yii::app()->request->getPost('addnumber'), 1, 100);
1101 $tokenattributefieldnames = getAttributeFieldNames($iSurveyId);
1102 $i = 1;
1103
1104 for ($b = 0; $b < $number2add; $b++)
1105 {
1106     while (in_array('attribute_' . $i, $tokenattributefieldnames) !== false)
1107     {
1108         $i++;
1109     }
1110     $tokenattributefieldnames[] = 'attribute_' . $i;
1111     Yii::app()->db->createCommand(Yii::app()->db->getSchema()->addColumn('{{tokens_'.intval($iSurveyId).'}}', 'attribute_' . $i, 'VARCHAR(500)')->execute();
1112     $fields['attribute_' . $i] = array('type' => 'VARCHAR', 'constraint' => '500');
1113 }
1114
1115 LimeExpressionManager::SetDirtyFlag(); // so that knows that token tables have changed
1116
1117 Yii::app()->session['flashmessage'] = sprintf($clang->gT("%s field(s) were successfully added."), $number2add);
1118 Yii::app()->getController()->redirect(Yii::app()->getController()->createUrl("/admin/tokens/sa/maagetokenattributes/surveyid/$iSurveyId"));
1119
1120
1121

```

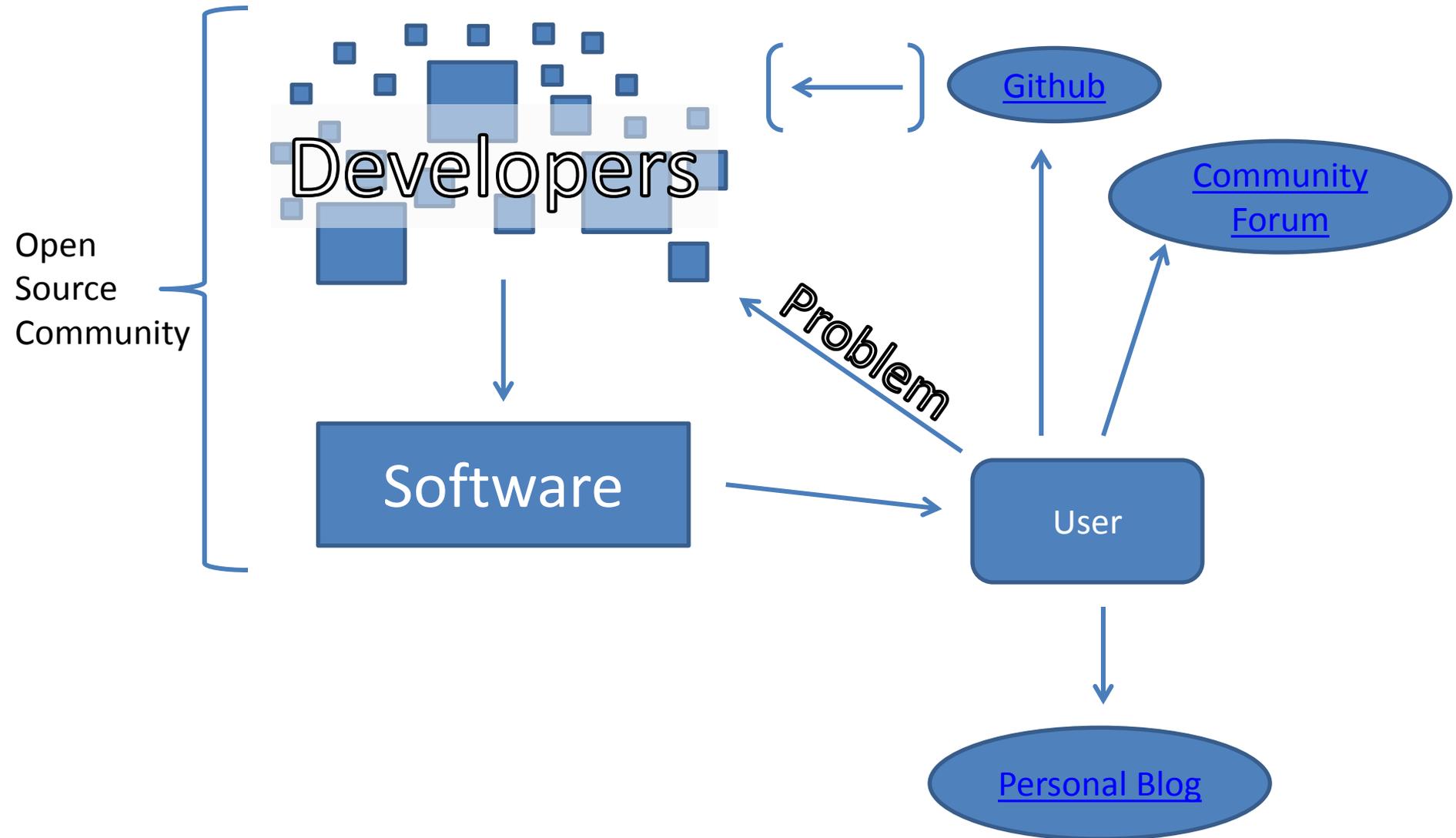
Find result - 4459 hits

Search "255" (4459 hits in 244 files)

- C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\export.php (4 hits)
- C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\remotecontrol.php (3 hits)
- C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\tokens.php (2 hits)
- Line 1115: Yii::app()->db->createCommand(Yii::app()->db->getSchema()->addColumn('{{tokens_'.intval(\$iSurveyId).'}}', 'attribute_' . \$i, 'VARCHAR(255)')->execute();
- Line 1116: \$fields['attribute_' . \$i] = array('type' => 'VARCHAR', 'constraint' => '255');
- C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\VerificationController.php (5 hits)
- C:\Users\Andrew\Desktop\limesurvey200plus-build130526\limesurvey\application\controllers\admin\statistics\helpers.php (8 hits)

PHP Hypertext Preprocessor file length:116630 lines:2404 Ln:1111 Col:150 Sel:0 UNIX ANSI INS

A path for open source work



Some final thoughts

- Breaking stuff helps you learn and connects you to other people with the same issues
- Building it yourself gives you unique tools for no \$\$\$ and often results in broadly applicable tools
- Time and frustration are the inevitable costs of such projects
- Sometimes, it is the best (only) option

Make it with friends

Break it yourself

Open-source software and DH research

Andrew J. Roback

PhD Candidate, IIT

Acknowledgements

Break It Yourself is the title of an album by Mom+POP records' recording artist Andrew Bird. I heard the title on WXRT (93.1FM) Chicago and it partially inspired the title of this talk.

Thanks to my academic advisor, Libby Hemphill, for recommending to me the technologies discussed in this presentation. Thank you.

Thanks to Karl Stolley, faculty member in the Humanities department, for helping me better understand the open-source initiative and associated technologies.

Thanks for financial support from the IIT Humanities Department and the National Science Foundation.