Tool: Technical guide

Utilising the data

Ian Field

Table of Contents

Introduction 2

Setup 3

Locales 3

Database 3

Data Use 4

Data Structure 4

File Structure 5

# Introduction

This tool has been created to be a flexible platform for gathering result sets from public participation of content. For example, a block of text can be presented on the web page, beneath that a series of check boxes will be shown with values associated with the text. This could be used as an assessment platform to analyze human input or reaction.

The tool supports multiple languages with a few elements of user customization throughout. By editing the text-based files it is possible to provide a limited customization of the site.

# Setup

This tool is designed to run on a web server, with requirements of PHP and MySQL in order to function. The settings for the database connection are set within PHP variables located in the *html > db\_conn* folder within the file ‘*config.php*’. The $dbhost variable should contain the server address and port number appended to the end in the following format: “:####” where required.

## Locales

The tool supports language customization to meet your target audience’s needs. These are altered in a similar manner to above. The content of the ‘About’ page can be specified within *html > config > about.php*. This is essentially a text file however, feel free to add any extra HTML tags you require for your desired display.

Within the same folder is another file with a similar format to the database configuration file. Here you may customize the about page title and navigation link text.

Below is an example set up for English:

<?php

$link\_about = "About";

$link\_home = "Home";

$about\_page\_title = "About this site";

?>

## Database

To set up the database to be ready for use locate the file DB > setup.sql. This should then be executed on the database server to serve as an all-in-one setup process.

Next visit your website and you can begin using it straight away with the first pre-populated database entries.

# Data Use

The database does not store user information past that of their IP address, which can later be used to remove spam or erroneous entries within your database.

During the design of the site it was decided that to maximize the likelihood of repeated, honest participants in the survey or other data gathering. With this reduces the possibilities of filtering spam or malicious entries through the use of something like a capcha ( "**C**ompletely **A**utomated **P**ublic **T**uring test to tell **C**omputers and **H**umans **A**part”), which essentially would present the user with an obscured word to type in. This was felt to delay the process or rapid result gathering that the site is intended for so the database results should be used with caution.

It is hoped that the database ‘entries’ table is not altered during a data collection phase, and is then re-built upon a change. This is due to the nature of the randomization procedure on creating the data capture page. Should a Primary Key index no longer exist it is expected that the page would produce an error message rather than a randomized entry.

Care has been taken to ensure that despite displaying the options in alphabetical order with no duplicates, they still correctly correspond to the same locations within the database should your option set require this.

The database is structured to allow three option groups. A, b, and c with each group holding 5 options. The ordering of these is not reflected on the website representation of the data as it is re-ordered to be alphabetical.

## Data Structure

The data structure and relation is the following:

Entries(**entry\_id**, corpus\_title, corpus\_body, a1, a2, a3, a4, a5, b1, b2, b3, b4, b5, c1, c2, c3, c4, c5)

Selections(**selection\_id**, entry\_id, user\_ip, selection\_time, a1, a2, a3, a4, a5, b1, b2, b3, b4, b5, c1, c2, c3, c4, c5)

The corpus title in the example files has a limited field size of 100, the body has a limit of 2500 characters and each of the options has a 100 character limit. These can be changed as appropriate either through an administration tool or within the SQL script. For more detailed information on the fields in the tables please see DB > create.sql.

The Selections table is designed to require the least space, and allow for very fast queries due to operating on a Boolean data type as it is expected to have a large number of rows over the experimental period.

The option groups in the selections table are Booleans, which are intended to directly correspond to the entries table options. The selection time and the user\_ip address are automatically populated to facilitate the customized filtering of the result set should a particular ip\_address or time slot be identified as spam when mining the data.

There is a relation between the two tables via the entry\_id key for more complex SQL queries. Both entry\_id and selection\_id are automatically incremented, and as such do not require calculation upon an INSERT function.

Within the folder named DB with some example SQL scripts. Including adding and removing constraints on tables, along with a single script to set up an example database with two entries populated.

# File Structure

This section details the construction of the web pages and the location of various site components. The DB folder contains the SQL files needed to set up the database. The file names are fairly self-explanatory. The Docs folder contains this document and a user manual. The html folder is the contents of the site for use within the public html folder of the webserver.

Within the html folder are:

|  |  |
| --- | --- |
| Index.php | The initial entry point for the website |
| Navigation\_bar.php | The nav bar links, included on all pages |
| Process\_entry.php | The form handler, which processes the POST from javascript and then updates the database with the user’s options |
| Footer.php | The footer which is included on each page of the website |
| About/index.php | The about page for the website |
| Config/about.php | The text file to populate the about page with. |
| Config/locales.php | The file which contains the php variables for the link text and the home page title text. This is to support a different language, however it is assumed that the text used will be html compliant (e.g. Romanised). |
| Content/privacy.php | The privacy policy should it be required. |
| Content/terms.php | The terms and conditions page, should it be required. |
| Css/reset.css | The reset stylesheet to hopefully reduce cross-site scripting. |
| Css/style.css | The actual stylesheet used for the web page. |
| Db\_conn/closedb.php | Close the opened database collection, just used for a shortcut. |
| Db\_conn/config.php | The configuration file for the database connection, including the database, username, password and location. |
| Db\_conn/opendb.php | Open the database connection specified within the configuration file. Used for a shortcut in the php code. |
| Javascript/form.js | This is the main event handler and posting handler of the database. It utilises JQuery to perform the POST operation for the form.  The submission is prevented should no user interaction be made to the page, this is to reduce the number of spam results for later analysis.  The selection choices are maintained even with the selection of ‘none of the above’, however they are not updated in the database unless ‘select from below’ is selected. |

Each file has been commented or utilises clear coding syntax to ensure modification of the tool is simple and easy.