mozilla Labs

TEST PILOT

Aggregated Data Samples

Info

- Test name: A Week in the Life of a Browser – Version 2.
- Test duration: 7 days
- Test type: Global
- Firefox versions covered: Fx 3.5 and Fx 3.6, Fx4 Beta
- Data submission: 527,817 test sets submitted in November 2010.

Data

Anyone is welcome to download and use the aggregated data files below.

Filename	Download Size	Num. of Tables	Num. of Users	Description
witl.db.gz	1.1 GB	3	About 27,000	Gzipped SQL dump of all three tables
witl_large.tar.gz	469 MB	3	About 27,000	Gzipped Tar archive of three CSV files, one for each table
witl_small.tar.gz	7.4 MB	3	About 27,000, 387 w/ event data	Gzipped Tar archive of three CSV files, one for each table

All of this data has been collected and is being shared under the terms of the Test Pilot Privacy Policy from Mozilla Labs. We license this data for use by anyone under the terms of the Creative Commons Attribution 3.0 United States License.. This license allows derived works as well as commercial use, as long as you give attribution for the source of the data.

There is no personally-identifiable information, and no URLs, contained in this data set.

Detailed Schema

There are two ways to work with this data sample: you can download the SQLite database, which consists of 3 tables, or if you are more comfortable working with .csv's, you can download the witl.tar.gz file, which contains the same 3 tables as separate .csv's. See below for a description of the tables and their columns.

Table Name: 'users'

Main table of users and the metadata associated with each user. Note: 'user_id' is consistent over all tables - tables can by merged by/on 'user_id'

Column Name	Explanation
user_id	All rows that have the same user number came from the same submission. Each submission is assumed to be a different user, but there are various ways that a single user could make multiple submissions (e.g. multiple computers, multiple Firefox profiles, etc.) These user numbers have no meaning or correlation to anything outside of the aggregated data file in which they appear. Their order has been randomized, so the numbers do not correspond to the order in which submissions were received.
location	Not used
fx_version	Firefox version
os	OS version
version	Test Pilot Extension version
survey_answers	Not used
number_extensions	Number of installed add-ons/extensions

Table Name: 'events'

Main table of users and recorded events.

Column Name	Explanation
user_id	See 'users' table'
event_code	Integer defining what the event was; see below for a description of these event codes.
data1	String; interpretation depends on event_code
data2	String; interpretation depends on event_code
data3	String; interpretation depends on event_code
timestamp	The time at which the event occurred, given in milliseconds since epoch
session_id	Not used

Event Codes

event_code number	Name	Explanation	Meaning of data1, data2, data3 columns
	STUDY-STATUS		'data1' identifies the version number of

0	STUDY_STATUS	Recorded on study startup.	'data1' identifies the version number of the study (currently 5).
	BROWSER_START	Recorded on application startup.	No additional data recorded.
2	BROWSER_SHUTDOWN	Recorded on application shutdown.	No additional data recorded.
3	BROWSER_RESTART	Recorded on application shutdown, if the user triggered the shutdown by clicking "Restart Firefox" (for instance in the add-ons manager).	No additional data recorded.
4	BROWSER_ACTIVATE	Recorded when the browser returns to active use after going idle.	'data1' column identifies the mechanism used to detect activation (basically debugging info): "Self-ping timer" or "IdleService observer".
5	BROWSER_INACTIVE	Recorded when we detect that the browser has gone idle, either through the self-ping timer (study pings itself every five minutes while firefox is active) or through the IdleService observer (which triggers after ten minutes of Firefox being idle).	Data1 column says either "Self-ping timer" or "IdleService observer".
8	BOOKMARK_STATUS	Recorded on app startup.	'data1' is the total number of bookmarks found. 'data2' is the total number of bookmark folders. 'data3' is the maximum depth of bookmark folders.
9	BOOKMARK_CREATE	Recorded when user creates a new bookmark OR a new bookmark folder. Ignores Livemark folders and does not trigger when a bookmark is created in one of these.	'data1' is either "New Bookmark Added" or "New Bookmark Folder".
10	BOOKMARK_CHOOSE	Recorded when user chooses a bookmark.	No additional data.
11	BOOKMARK_MODIFY	Recorded when user deletes a bookmark or moves it to a different folder. Ignores Livemark folders.	Data1 is either "Bookmark Removed" or "Bookmark Moved".
12	DOWNLOAD	Recored when a download completes.	No additional data.
17	PRIVATE_ON	Recorded when user enters private browsing mode. When in private browsing mode, no further recording is done.	No additional data.
18	PRIVATE_OFF	Recorded when user exits private browsing mode.	No additional data.
19	MEMORY_USAGE	Memory usage is recorded once every 15 minutes (although this was originally set to 1hr so data prior to Nov 6 will be 1 hour intervals), and upon startup. When memory usage is recorded, it will iterate through all the reporters in the memory manager, recording a MEMORY_USAGE event for each reporter.	'data1' contains the reporter path, and 'data2' contains the amount of memory used, in bytes.
20	SESSION_ON_RESTORE	Recorded on application startup, when session restore triggers.	'data1' is the number of windows restored, 'data2' is the number of tabs restored across all windows. If no session restore occurs, will record 0 for both columns.
22	PLUGIN_VERSION	On experiment startup, experiment records one PLUGIN_VERSION event for each plugin.	'data1' is plugin filename, 'data2' is plugin version, 'data3' is plugin name.
23	HISTORY_STATUS	Recorded on experiment startup.	'data1' is the total number of distinct places in the Places database, i.e. history items.
24	PROFILE_AGE	Recorded on experiment startup.	'data1' is the age (as a timestamp) of the oldest file in the profile directory. (Using the last-modified date, since creation date is not always available)
25	SESSION_RESTORE_PREFERENCES	On experiment startup, recorded once for each preference relevant to session restore.	'data1' is name of the pref, 'data2' is the value. Recorded prefs are browser.sessionstore.resume_from_crash, .resume_session_once, and .max_resumed_crashes.
26	NUM_TABS	Recorded whenever memory is recorded (at startup and every 15 minutes) and also whenever a window or tab is opened or closed.	'data1' is number of windows, 'data2' is number of tabs (including the one just opened).
27	STARTUP_TIME	Recorded on startup. Will be used in the future to record time taken for Firefox to start up; however this data is not yet available.	

Table name: 'survey'

Table of users and associated Beta Background Survey answers. For each row, the first column is 'user_id' and each of the following 14 columns represent the 14 survey questions. Answer choices are coded to integers, so for a given question, '0' represents the first answer choice, '1' corresponds to the second answer choice, and so on. On some questions, users were allowed to check more than one answer. If they did so, the column will contain the numbers of all checked answers, delimited with pipe ("|") characters. Note that because the survey was optional, the table does not include all users.

Survey Question	Answer 0	Answer 1	Answer 2	Answer 3	Answer 4	Answer 5	Answer 6	Answer 7	Answer 8	Answer 9	Answer 10
How long have you used Firefox?	Less than 3 months	3 to 6 months	6 months to a year	1 to 2 years	2 to 3 years	3 to 5 years	More than 5 years				
Do you use more than one browser in your daily life?	No. I only use Firefox.	Yes, I use other browsers besides Firefox.									
If you use other browsers besides Firefox, what are they?	Chrome	Safari	Opera	Internet Explorer (all versions)							
If you use multiple browsers, what do you consider to be your primary browser?	I don't use other browsers, only use Firefox	Firefox	Chrome	Safari	Opera	Internet Explorer (all versions)					
What is your gender?	Male	Female									
How old are you?	Under 18	18-25	26-35	36-45	46-55	Older than 55					
How much time do you spend on the	Less than 1 hour	1-2 hours	2-4 hours	4-6 hours	6-8 hours	8-10 hours	More than 10 hours				

Web each day?	1 hour	hours				hours	hours				
How would you describe your computer/web skill level?	(Scale 1-10)										
Where do you usually access the Internet?	Home	Work	School	Mobile							
What smartphone device are you currently using?	Apple iPhone	Google Android (ex: Droid, HTC Hero)	Microsoft Windows Mobile (ex: HTC Imagio, Samsung Omnia)	Nokia Maemo (ex: N900, N810)	Nokia Symbian (ex: N97, E71)	Palm (ex: Pre, Treo)	RIM Blackberry	I don't use a smartphone.			
What are the main reason you use the Web?	Work: coding related	Work: non-coding related	School	Personal life assistance	Communication	Socializing	Entertainment				
What are your most frequently visited websites?	Search engines	Video sites	News sites	Social networking sites	Browser based games	Shopping	Paying bills and online banking	File download sites	Webmail	Forums	Adult pages	Gambling and online betting
How do you find out about the latest computer and technology news?	Online news sources: Google News, CNN,etc.	Online blogs: tech blogs, news blogs, etc.	Offline news sources: printed newspapers, magazines,etc.	Someone you know: a coworker, friend or family member,etc.	Advertisements	Social media: Facebook, Twitter, Ning, etc.						
How would you describe yourself as a web user?	(Free entry)										

Filtered Extensions

The following extensions are known to cause major changes to the relationship between tab use and memory in Firefox. Data submissions from users who had one of these extensions installed have been filtered out of the data sets provided on this page, to ensure that the data sets are as representative as possible of default Firefox behavior:

- Bar Tab