Creating 'Web Searches' for use in Reunion® 10

By Keith Bage, Reunion User

Revision 1.0 – 3 July 2012

Overview

This document is intended to provide information on how to create additional Web Searches to be used with Reunion 10's excellent **'Search the Web'** functionality.

It is not intended to detail how to use Reunion 10's 'Search the Web' functionality itself or to replace any of Reunion 10's excellent user manual. Rather, it is a supplemental document of a more technical nature on specifically how to create and add new web searches for use in Reunion 10. For information on how to use the standard Search the Web feature please refer to the Reunion 10 user manual under "**Searching the Web for more information**".

This document is the sole work of Keith Bage and is in no way associated with Leister Productions Inc. (the makers of Reunion 10), though this could not have been created without their kind help and encouragement.

I am happy to offer any assistance I can (where possible), but please keep in mind that time as always is limited, and this document does not imply any support.

What is a Web Search?

If you've stumbled across this document I'd say that there is a fair chance you already know what a Web Search is in the context of Reunion, but for those who don't here goes.

A Web Search provides a mechanism to perform an instant search of a specific web site for an individual in Reunion using the information recorded for that individual as search criteria.

A word about Websites

Before we move on to how to create a Web Search it is important to understand that it is not possible to create a web search for all websites due to the ways in which some websites have developed their search engines.

When investigating a new website, carry out a search by hand and take a close look at the URL of the results page. If the URL contains some of your search criteria as part of the URL then it is likely that you can create a web search for this website.

For instance: If I were to carry out the most basic of searches on Ancestry.com for the first name **Keith** and the surname **Bage** the resulting URL will look something like the one below.

http://search.ancestry.com/cgi-bin/sse.dll?gl=ROOT_CATEGORY&rank=1&new=1&so=1&MSAV=0&msT=1&gss=ms_f-2_s&gsfn=**Keith**&gsln=**Bage**&uidh=xu4

Looking closely at the URL you can see that it contains the words **Keith** and **Bage** and is therefore a good contender for the creation of a web search (a web search for Ancestry.com was provided as standard with the release of Reunion 10 and is mentioned here for illustration purpose only).

Where are Web Searches Stored?

Each web search is created as a special file with a **.plist** extension. These file types are actually Property Lists and are often used for storing various settings used by programs on the Mac.

These files are located under the following path, where **'YourUserDirectory"** is the name of your personal user folder.

'YourUserDirectory' > Library > Application Support > Reunion > Search Engines

By default Apple have hidden the user **'Library'** folder as of the release of Lion. To locate this folder from the Finder, hold down the alt key and from the menu choose **Go** and select **Library**.

TIP: For those familiar with using the Terminal, the Library folder can be made permanently visible by entering the following in the Terminal.

chflags nohidden ~/Library

Creating a .plist File

A .plist file is essentially an XML file (a specially formatted text file) and can be easily created or edited by using any basic text editor such as TextEdit. Other specialist text editors typically designed for programmers or web developers can also be used, such as BBedit or TextMate.

Strangely, possibly the easiest tool to use is Apple's Xcode developer tools which is free to all users. I say strange because the Xcode toolset is a complex set of tools that are really aimed at software developers for Apples OSX and IOS platforms. However, Xcode's plist editor strips away the need to see the pure XML code and simplifies input. Xcode is free in the Mac App Store or you may find a copy on your Mac OSX install DVD which you received with your Mac.

For the purposes of this document I will assume that users are using a text editor to create/edit the plists.

I find it easy to simply duplicate and rename an existing web search within the finder and then open it in my chosen text editor or in Xcode. By duplicating a file I typically only need to change a small portion of the file which determines the resulting URL.

plist File Structure

A .plist (Property List) file is largely made up of "keys" and "strings". The key's identify the element of data which is used either in Reunion's web preferences or is passed in the creation of the URL. The strings identify the values that apply to these keys.

The available keys and possible values can be seen in the following table.

Keys	Strings	Comment
EngName	'Any text'	A Description of the web search i.e. "***British Newspaper Archive***". Appears in your "Search the Web" All sites and user groups lists.
subtitle	'Any text'	Web address of the website for which the web search is being created. Appears in the preferences window for the selected site.
description	'Any text'	Information about the website. Usually a simplified version of the websites 'About' text. A comment may be added in square brackets i.e. **[Requires PAID subscription]**.
homepage	Websites URL	This is the URL of the page to be taken to when clicking the "Visit

		this site's Home Page" in the preferences window
context	'resource v1'	If Reunion does not see this key and value the web search will be ignored when Reunion is launched and will result in the web search not being available for use.
uid		Currently not used, but this is effectively a way of giving the web search a unique number. Numbers 1 – 1024 are reserved for Leister Productions Inc. This key can be omitted when creating your own web searches.
~~selected~~	~~true or false~~	~~Defines whether the web search is selected by default in Reunion~~ This key is no longer used as of Reunion 10.0.3
shouldEscape	true or false	Almost always true
spc	+	This is the character used by the search engine for spaces. Whilst I've found that this is typically '+' this can differ from site to site.
urlt	?	This is the first character that follows the initial search part of the URL. Again, whilst I have found '?' to be the most commonly used character, this can differ site to site.
varsep	&	This is the value used by the search engine between variables. Again, this can differ site to site.
url	"array"	This is a built array that will make up the URL to be passed to your web browser (see the 'Building the URL key' section below)
sex	"array"	Used to determine the value for the sex of the person used in the URL.

Building the URL Key

The URL key mentioned in the previous table is made up of an array of additional strings of information. These strings represent the different elements of the resulting URL. In turn these elements are made up of information from different "tags" which relate to information held within specific fields in your Reunion database. The table below shows the various tags that can be used.

Tag	Use	Comment
[GN]	Given Name	
[SN]	Surname	
[FNO]	First Name Only	
[FMLN]	First Middle Last Name	
[FNI]	First Name Initial	
[BPL]	Birth Place	
[DPL]	Death Place	
[BDATE]	Birth Date	Must pass a format with this tag (see below)
[MDATE	Marriage Date of the preferred marriage	Must pass a format with this tag (see below)
[DDATE]	Death Date	Must pass a format with this tag (see below)
[ADATE]	Any Date	Must pass a format with this tag (see below)
[SX]	Sex	
[i]	Additional URL strings	These are pieces of information used by the search engine which do not come from your Reunion fields

These tags need to be used in the <array> section of the URL key to effectively build the URL piece by piece. Due to each search engine using its own URL structure it is impossible to detail how to create the URL array. You will need to perform some manual searches for the website you are trying to create a search for and try and build up an equivalent array by breaking the URL into component parts between each occurrence of the '&' symbol (note though that some website may use a different character than '&' to separate variables).

For instance, lets say I want to create a search for nationalarchives.gov.uk. The link below illustrates a search for keywords **John Bage** with a from date of **1800** and an end date of **1900**.

http://nationalarchives.gov.uk/search/search_results.aspx?Page=1&QueryText=John+Bage&DateFrom=1800&DateTo=1900&SelectedDatabases=A2A%7cARCHON%7cBOOKSHOP%7cCABPAPERS%7cDOCUMENTSONLINE%7cEROL%7cMOVINGHERE%7cNRA%7cNRALISTS%7cPREM19%7cRESEARCHGUIDES%7cE179%7cCATALOGUE%7cWEBSITE%7cTRAFALGAR&SearchType=Quick

The first 'string" of the array would be the initial search prior to the ? symbol.

<string>http://nationalarchives.gov.uk/search/search_results.aspx</string>

Not all pieces of the URL are essential and can be omitted but this is really a case of trial and error. In this case we don't need the **Page=1** element and so the next component we need is the one which is **QueryText=**. In this example I'll use the **[FMLN]** tag so as to pass in the individuals full name. Therefore the next <string> element would be

<string>[FMLN]QueryText= %@</string>

The **%@** symbols are used to represent the content from your Reunion fields.

Next we need to bring in the "from date" and for this I'll use the **[BDATE]** (birth date) tag with the format set to just include the year, which results in the following string.

<string>[BDATE-YYYY]DateFrom=%@</string>

Finally I want to include the "to Date" and for this I use the **[DDATE]** tag, again with the Year (YYYY) format to provide a string of:

<string>[DDATE-YYYY]DateTo=%@</string>

So, for this example that completes the <array> of the URL. I am not going to include the SelectedDatabases section of the URL as it doesn't appear to be needed (many search engines seem to make certain assumptions when data is omitted from the URL).

So, the complete url key and <array> is as follows:

```
<key>url</key>
<array>

<string>http://nationalarchives.gov.uk/search/search_results.aspx</string>
        <string>[FMLN]QueryText= %@</string>
        <string>[BDATE-YYYY]DateFrom=%@</string>
        <string>[DDATE-YYYY]DateTo=%@</string>
</array>
```

A special word about using Date Tags

Date type Tags must also include a format for the date. Available formats are:

DDMMYYYY = 05/02/1981
MMDDYYYY = 02/05/1981
DDMonYYYY = 05 Feb 1981
YYYY = 1981 (aka Year Only)

These formats must be passed with the date tag i.e. [BDATE-DDMMYYY]

Another special word on the sex array

Most websites do not use a full description of the persons sex in the URL. Instead they use abbreviations such as 'f' or 'm'. The sex array associates an appropriate value with the persons sex. This is represented in the .plist file as a special 'dictionary' (key-value pairs).

So a typical sex key array would be as follows.

```
<key>sex</key>
<dict>
        <key>fml</key>
        <string>m</string>
        <key>ml</key>
        <string>f</string>
        <key>mt</key>
        <string></string>
</dict>
```

Testing the Web Search

You will need to periodically test your web search by saving the .plist file and re-launching Reunion 10 (note that Reunion will need to be restarted to take note of any changes to web searches) then trying the search for an individual from your database.

If you find that your web search does no appear in Reunion, then it is likely that either:

    a. The .plist file does not contain a correct 'context' key
    b. The syntax of the .plist file is incorrect
    c. The file is not in the correct directory.

A word about Ordering Your Searches

Reunion will alphabetically sort all Web Searches contained in the **All Sites** list based upon the value contained in the EngName key within the Web Search.

Please note that prior to Reunion 10.0.3 this sorting was based upon the name of the .plist file itself.

Wrap Up

I hope that you have found this document to be of use. Because of the variable nature of websites and their search mechanisms it is a difficult topic to put into writing. Hopefully though what you have here is a good starting point in helping you understand how Reunion's web searches are constructed.

I would like to thank Leister Productions Inc. for all the information they have freely given me and for being encouraged by my enthusiasm for web searches and for making significant improvements in this area in Reunion 10.0.3. Especially in light of how quickly these come after Reunion 10's intial release.

You can contact me at keith@bage.org.uk or at my website www.bage.org.uk where I research any occurrence of the name BAGE worldwide. I can also be found as @keithbage on Twitter or of course on the ReunionTalk forum at http://www.reuniontalk.com.