

Searching and Sorting Tibetan fields in Microsoft SQL Server

A brief guide by Robert Chilton (ACIP)

A. Client computer configuration (clients running Mac OS 10.5 or later can also be used)

1. Install any desired Unicode Tibetan fonts.
2. Enable complex scripts in Windows XP [this step is not needed in Vista / Windows 7]. This step usually requires access to the \I386 folder found, for example, on the Windows XP installation CD.
3. Install the Tibetan Keyboard and, in Vista / Windows 7, the Tibetan Locale. In Windows XP, associate the keyboard with an Input Language (such as Tswana).
4. In the case of Windows XP [this step is not needed in Vista / Windows 7], all the programs that use and display unicode Tibetan script must have access to a Tibetan-capable version of Microsoft's 'Unicode Script Processor', also called 'Uniscribe', which is in a file named 'usp10.dll'. The file usp10.dll must be placed in each program application folder, such as Internet Explorer, Mozilla Firefox, and any other client-side applications that will display Unicode Tibetan data.
5. In the web browsers, default font(s) for Tibetan (or 'Other Languages') should be set to a Unicode Tibetan font.
6. In the web browsers, the default character encoding should be set to 'Unicode (UTF-8)'.

B. Enabling Microsoft SQL Server to return matching results of searches on Tibetan fields in Tibetan alphabetical order

You must use SQL Server 2008 (or later) to enable search matches to display in Tibetan alphabetical order (according to Microsoft collation order which is about 95% correct for Tibetan).

Here are the steps:

1. Unless the database is very large, it might be possible to use the free (no-cost) version of SQL Server 2008 known as SQL Server 2008 Express. You should carefully check the limitations and requirements to see if the Express version of SQL Server 2008 will work in your situation. After downloading and beginning the install process you may find you need some other prerequisites such as Windows Power Shell and others items. These can be downloaded and installed; the SQL Server install program tells you which items you need.
2. You should get and install the version with tools because that gives you SQL Server 2008 Management Studio which you need (the earlier versions won't work with SQL Server 2008). All of this will run on Windows 2003 Server or Windows 2008 Server or even Windows XP. Also, there is an Import and Export tool that can be used to migrate the data out of a previous SQL server version into the new version.
3. In Management Studio you need to connect to your database, expand tables and then right click the table and select design to bring up the table's columns in design mode. Now select the column you want to be able to sort in Tibetan. In the column properties window look for table designer. The entry just below this is Collation— usually set to Database Default. Highlight this and a button appears to the right. Click the button and the Collation window opens. Select Windows collation and in the drop down below this select Tibetan_100. There are some other settings here (binary sort, BIN2, dictionary sort...) that you can select according to your desired collation system. Click OK and close the collation window.
4. In terms of how you use the SQL database to feed the web pages, this will depend on whether you are using, e.g., Classic Active Server Pages (ASP) or else ASP.NET. If you use Classic Active Server Pages (ASP) there are three things that need to be looked at to get Tibetan to display properly:
 - a. You definitely need to make sure the ASP files are saved as UTF-8 files. You can do this simply by loading each file into Notebook (they're simple text files) and then re-saving them using Save as... Use the file type drop down to change the file type to UTF-8. By default you'll find that files are saved as ANSI (which is a big problem for displaying Unicode Tibetan data!).

b. As an example, you can use the following header format for each page that involves displaying HTML:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<html xmlns="http://www.w3.org/1999/xhtml">
<head>
  <title>Some page title</title>
  <meta http-equiv="Content-Type" content="text/html; charset=utf-8"
/>
  <link rel="stylesheet" type="text/css" href="some.css">
  <script language="JavaScript" type="text/javascript"
src="some_functions.js"></script>
</head>
```

Obviously the page <title> and .css file and .js file will vary according to what you need for the specific pages.

c. Finally, use the server.HTMLEncode method for display. For details of use and code examples see:

<http://msdn.microsoft.com/en-us/library/ms525347.aspx>

5. Here is an example SQL search query for giving results for a partial match of the search string with results sorted in Tibetan alphabetical order by a chosen Tibetan field (column):

```
Select <field_1>, <field_2>, ... <field_n> from <table> where <searched_on_Tibetan_field> like N'%<unicode_Tibetan_search_string>%' ORDER BY
<Tibetan_field_to_use_for_ordering_results_in_alphabetical_order>
```

For exact matching, use operator '=' instead of 'like'.