

DYNAMIC WEB PAGE WITH CONVERSION PROCESS FOR CONVERTING DATABASE FIELDS FROM ONE INDIAN LANGUAGE TO OTHER INDIAN LANGUAGE

Dr. Dhiren Patel

ABSTRACT

E-government contains huge data volume, if we store all Indian languages records then again volume increase. After coming Unicode Indian languages conversion is possible. Also .NET and JAVA technology internationalization is power-full future for working in multilingual environment. Depending on client's request use of language conversion is possible at client side by executing script code. System take care that culture of current environment language and data record's language are different than only perform mathematical conversion. On second phase verify that converted Unicode value has assigned any character or not.

Keywords: E-government, conversion process, Unicode, ASP.NET, C#.NET, Culture

1. INTRODUCTION

In India, e-government projects are increasing day by day so as to facilitate more services to citizens. India is spread in 28 states and 7 Union territories, each state and Union territories has a unique demography, history and culture, dress, festivals, language^[5]. India has major spoken languages are Hindi, Punjabi, Tamil, Marathi, Telugu, Gujarati, Kannada, Assamese etc. Hindi speaking persons are 18,20,00,000, Punjabi speaking persons are 7,83,00,000, Tamil speaking persons are 7,50,00,000, Marathi speaking persons are 6,81,00,000, Telugu speaking persons are 6,98,00,000, Gujarati speaking persons are 4,65,00,000, Kannada speaking persons are 3,54,00,000, Assamese speaking persons are 1,53,00,000. Majority spoken Indian language ^[7], belong linguistic family is Indo-European languages Indo-Aryan nearly 70% population in India [Error! Reference source not found.]. More than 65% population of India is leaving in rural area. Indian rural area literacy ratio is very poor. Government is announcing scheme for uplift rural citizen. Government wants to provide information of policy; scheme, land record and collection of tax through E-Government in citizens native language. In India many number of language with large number of native language speakers, also any citizen can easily move them self between any states of India. Here citizens require E-Government facility in their own native language.

2. UNICODE

Unicode Standard assigns each character a unique numeric value and name, no matter what the platform, no matter what the program, no matter what the computer scripting language. The Unicode Standard Version 5.2.0 provides the capacity to encode all of the characters used for the written language of the

world. The Unicode Standard follows a set of fundamental principles are “Universal repertoire”, “Efficiency”, “Characters, not glyphs”, “Semantics”, “Plain Text”, “Unification”, “Dynamic composition”, “Stability”, “Logical order”, “Convertibility”. Indian languages like as Hindi, Gujarat and Punjabi logical order and convertibility is same. Meaning is that first character in Hindi “ॐ” has Unicode value in Devanagari U-0905₁₆ (fifth character in Unicode Devanagari Script); first character in Gujarati is “ૐ” has Unicode value U-0A85 (fifth character in Unicode Gujarati Script). It means to convert from Hindi to Gujarati just change range of one language to other language range.

Hindi characters are included in Unicode Devnagri character set. It will start from value U-0900₁₆ to U-097F₁₆ its value in decimal system is 2304₁₀ to 2431₁₀. Unicode Punjabi character is known as Gurmukhi script character and it will start from value U-0A00₁₆ to U-0A7F₁₆ its value in decimal system is 2560₁₀ to 2687₁₀. Unicode Gujarati character will start from value U-0A80₁₆ to U-0AFF₁₆ its value in decimal system is 2688₁₀ to 2815₁₀. From above information if data is in Hindi and citizen want to see same data in Punjabi then that data’s character is in range of 2304₁₀ to 2431₁₀ will be converted in to range of 2560₁₀ to 2687₁₀, for this type of conversion in character value do addition of 384₁₀ by such process one character will be converted from one language to other language.

Language	Hindi	Punjabi	Gujarati
Hindi		+256	+384
Punjabi	-256		+128
Gujarati	-384	-128	

Table 1. From one language to other language conversion value.

Programmatically doing addition or subtraction from character’s value, perform casting from character to integer. $l = (\text{int})c + \text{convert_val}$; here l is int type of variable, c is Unicode character (char type variable) and convert_val is int type of variable has value for converting one language to other language conversion value. After doing that, again convert integer value in to character type then conversion from one language to other is over. For doing that again use casting from integer to character. $c = (\text{char})l$; This process is applicable to database filed language character set only not for other language, e.g. database filed is in Gujarati but in between English character is used, now in this case converting Gujarati to Hindi don’t convert English character and live it as it is.

3. CULTURE SETTING

For E-Government application, web technology is most appropriate available technology today. Many approaches are available for web technology; here

ASP.NET technology with C#.NET language for .NET 3.5 frame work technology is chosen. It is available with Microsoft Visual Studio 2008. To use Unicode character set of different Indian languages, multi-lingual facility we have to use. In multi-lingual facility one has to change culture type for new language/country. The Culture property is used to help localize page content. You can set it to any valid culture ID. For example, the gu-IN culture ID sets the page to Indian Gujarati^[11]. There are two types of cultures namely Culture and UICulture. Both are member of class System.Threading.Thread.CurrentThread so in program they are written as System.Threading.Thread.CurrentThread.Culture and System.Threading.Thread.CurrentThread.UICulture.

Normally culture has given pre-define naming format is “language code” “-” “country code”. Here language code is written in small latter, and country code is written in capital latter. For Indian languages coding is as following.

Sr.No.	Name of Language	Code	Language code define
1	Gujarati	gu-IN	2
2	Hindi	hi-IN	1
3	Punjabi	pa-IN	3

Table 2. List of used Indian Languages with there culture code value ^[13] and language no. usede in program code

3.1 Resource file

Culture is bounding with resource file to get information in format of culture’s language. Resource file in Visual Studio 2008 is created by two ways. One Local Resources and other is Global Resources. Here Global Resources is created and use for internationalization. Global Resource in Visual Studio web project create “App_GlobalResources” folder, under that folder create Global Resources file name for example is “LocalizedText.resx”, and enter require field name and its values in English. Then copy same file and paste it in to same folder, then rename that file to “LocalizedText.gu-IN.resx”. Now open “LocalizedText.gu-IN.resx” file and modify value of each filed in to Gujarati value. So when from program changing culture to Gujarati that time as a resource file instead of “LocalizedText.resx” program bound resource file “LocalizedText.gu-IN.resx”.

After creating required resource files, our job is to bound form’s objects with resource file field name. Now open form file in source mode, then reach at required field entry, and make correction as following code.

```

<asp:Label ID="s_name_lab" runat="server"
    Text="<%$ Resources:LocalizedText, name_lab %>"
    meta:resourceKey="s_name_labResource1">
</asp:Label>

```

In Syntax after Text = write “<%\$” then “Resources:” after that write resource file name without any extension (for above example it is LocalizedText). Second parameter is resource file field name, then lastly writes “%>”. When programmatically changing culture, Resources automatically give affect. So in project if you adding more language resources in future form code modification are not required.

Now programmatically citizen has to choose their language from list of available languages. After selecting language control send to new page where that page culture is to be initialize by program. For that for example form name is “Localized.aspx” then its code is to be written in the file name “Localized.aspx.cs”. Before Page_Load() method InitializeCulture() is overloaded for setting citizens language as default culture.

4. STEPS FOR CONVERSION PROCESS

After changing Culture and UICulture as per Table 1 defining two dimension variable and initialize it by value as per Table1. Here variable is taken as “int [,] con_val;”, variable’s first subscription indicating database record’s language number and second subscription indicating citizen’s required language number. By program we have to check that converting character is in given language then only do conversion process. Otherwise in different language non-significance conversion process may happen that will harmful for conversion process. So define language range by variable name “int[,] range_lan;” this variable’s first subscription indicating language number define by your program, and second subscription indicating language lower limit by ‘0’ and upper limit by ‘1’. For database aspect no extra care has to be taken, but only thing is to be kept in mind that use only such future that will give accessibly of individual fields of record of database. Here database is taken as Microsoft Access database.

After getting record’s individual filed value, our target is that record’s language is same as citizen’s required language? If both are same then display field’s value as it is, else perform conversion process. Designing such application’s database must take care that each record not in the same language. At the time of database table design, one field must for indication record language. For conversion process get record’s language from field of record and convert that in to its corresponding number. And same way get current Culture code (Table2), after that convert

Culture code in to its corresponding number. After completing such preparation conversion process code has to be return.

4.1 First step of conversion process code

```
protected string conv(string t_txt)
{
    long l = 0;
    l = range_lan[data_lan,0] + 5; //2nd subscript 0=low
    char c = (char)l;
    string return_st = "";
    for (int i = 0; i < t_txt.Length; i++)
    {
        c = t_txt.ToCharArray()[i];
        if ((int)c > range_lan[data_lan, 0] &&
            (int)c < range_lan[data_lan, 1])
        {
            l = (int)c + con_val[data_lan,culture_lan] ;
            c = (char)l;
        }
        return_st = return_st + c;
    }
    return (return_st);
}
```

4.2 Second step of conversion process

After processing conversion process step above now, there is possibility that first language has character but above step conversion new language can not have character. For example in Hindi fourth character is 'ः' Unicode value is U-0904 after conversion process to Gujarati language fourth character's Unicode value is U-0A84, this Unicode has not assign any character. So in conversion process take care of such case. By two methods are possible. First method is to put blank space for such character. Second method is to put original character as it is at that passion. List of Hindi character but not Gujarati character is as follow.

ः U-0903	ऐ U-090E	औ U-0912	ऋ U-0929	ॠ U-0931	ऌ U-0934	ॡ U-0946	ॢ U-094A
क U-0958	ख U-0959	ग U-095A	ज U-095B	ड U-095C	ढ U-095D	फ U-095E	य U-095F
श U-097B	ष U-097C	?	ॱ U-097E	ॲ U-097F			

Table 3. List of Hindi character which cannot have conversion character in Gujarati

Similarly we can prepare one language to other language conversion not assign character list.

5. Screen

★ Favorites Untitled Page

નામ : ધીરેન બ. પટેલ

સરનામું : ૬, નવરંગ કોલોની

નવરંગપૂરા

અમદાવાદ

પીનકોડ : ૩૮૦૦૦૯

ભાષા બદલવાના પાના ઉપર જવા <=> 1

Figure 1 : Screen to Show Data and Label in Gujarati

★ Favorites Untitled Page

नाम : धीरेन ब. पटेल

पता : ६, नवरंग कोलोनी

नवरंगपूरा

अमदावाद

पीनकोड : ३८०००९

भाषा बदलने वाले पने पर जानेके लिए <=> 1

Figure 2 : Screen to Show above Figure 1 Data and Label in Hindi

The screenshot shows a web browser window with a single tab titled 'Untitled Page'. The browser's address bar contains the Punjabi text 'ਪਨਾਭਾਸ਼ਾ ਬਦਲਨ ਵਾਲੇ ਪੈਨ ਪਰ ਜਾਨੇਕੇ ਲੀਏ'. Below the address bar, there are several form fields:

- ਨਾਮ :** ਧੀਰੇਨ ਬ. ਪਟੇਲ
- ਪਤਾ :**
 - ੬, ਨਵਰੰਗ ਕੋਲੋਨੀ
 - ਨਵਰੰਗਪੁਰਾ
 - ਅਮਰਦਾਦ
- ਪੀਨਕੋਡ :** 380009

At the bottom of the form, there are navigation buttons: '<=' and '=>', and a page number '1'.

Figure 3 : Screen to Show above Figure 1 Data and Label in Punjabi

5. CONCLUSION

Use of above work is for language conversion of person name (first name, second name, and surname), address, city name, state name, country name type of data into similar sense character order type of language. It is done by machine (computer) only, no human interference require for such process. Here logical work exercise done for minimum process for conversion process. Above process will remove data redundancy, in terms of same data in different languages. Due to that multi-language web site maintenance is quite simple process.

For sentence type of data conversion process is not better solution as per language aspect. But in future conversion process will help full for translation process or conversion process is the first step for translation process.

References

1. D. Wujastyk: Transliteration of Devanagari, (1996)
2. Abduljaleel, N. and Larkey, L. S.: Statistical Transliteration for English-Arabic Cross Language Information Retrieval : ACM Transactions on Asian Language Information Proceeding, New Orleans, LA, USA, P.139-146(2003)
3. AL-ONAIZAN, Y. and KNIGHT, K.: Translating Named Entities Using Monolingual and Bilingual Resources.: Proceddings of the 40th Annual Meeting of the Association of Computational Linguistics. Association for Computational Linguistics, Philadelphia, PA, USA, P.400-408(2002)
4. Indian Languages <http://headlinesindia.mapsofindia.com/culture/india-languages.htm>
5. State and Union Territories – Know India http://india.gov.in/knowindia/state_uts.php

6. Devanagari – Wikipedia <http://en.wikipedia.org/wiki/Devanagari>
7. List of languages by number of native speakers – Wikipedia http://en.wikipedia.org/wiki/List_of_languages_by_number_of_native_speakers
8. What is Unicode? <http://www.unicode.org/standard/WhatIsUnicode.html>
9. Microsoft .NET Internationalization <http://msdn.microsoft.com/en-us/goglobal/bb688096.aspx>
10. Translate Hindi to Punjabi Text <http://h2p.learnpunjabi.org.in>
11. Page.Culture Property (System.Web.UI) <http://msdn.microsoft.com/en-us/library/system.web.ui.page.culture.aspx>
12. Proposed changes in Unicode Standards for Indic Scripts- Gujarati http://tdil.mit.gov.in/prop_uni/new_gujarati.pdf
13. Table of Language Culture Names, Codes, and ISO Values [http://msdn.microsoft.com/en-us/library/ee797784\(CS.20\).aspx](http://msdn.microsoft.com/en-us/library/ee797784(CS.20).aspx)