HYBRID METHOD TO DESIGN MULTI-LANGUAGE WEB SITES

Mihai Bîzoi, Ana-Maria Suduc and Gabriel Gorghiu

Valahia University of Targoviste, Unirii Bv., No. 18-22, Targoviste, Romania

Keywords: Web design, Multi-languages, Educational Comenius project 2.1.

Abstract: Language promotes the most suitable way for expressing ideas, habits and values. For a multinational educational project (Comenius 2.1.) with a widespread European target group, it is very important to provide the online educational resources (courses, virtual instruments, etc.) through the project website, in several languages: English (as a worldwide spoken language) and partners' languages. This paper briefly presents three methods to create multi-language web sites and proposes a new architecture – the hybrid method architecture. In addition an example of a multi-language web site implementation is illustrated.

1 INTRODUCTION

Communication is based on language even it involves a mixture of verbal and nonverbal messages or a blend of written and spoken messages. The huge development of telecommunications - including television, radio and the Internet - has led (in the last years) to increasingly sophisticated combinations of visual, spoken, and written "texts" (Tiscali, 2008).

The European Union, had, from the very beginning, as heart of its policies, the preservation of the linguistic diversity. The EU policy of multilingualism as a deliberate tool of government is unique in the world and the use of EU citizens' languages is one of the factors which make it more transparent, more legitimate and more efficient.

2 METHODS TO CREATE MULTI-LANGUAGE WEB SITES

In the *Consideration in the Design of Multi-language Websites* paper, Zev Itzkowitz and David Grunwald propose four architectures for a multi-language web site development. The basic ideas of these methods consist of common operations such as replication and dynamic content generation.

First method, *site replication* is the most common method used to create multi-language web sites. Virtually, the web site consists of architecture of directories, each directory containing the same web site, but in other language. One of these sites is the default site and to access it, in the root of the web, a file that generates a redirection is used. To change the default language, only the redirection file from the root site must be changed.

The European Union web site is the most representative one for the multi-language web site category, mainly due to the fact that it is the most translated site – at present, its translation is available in 23 languages. This web site is built based on the replication method but with different architecture of directories– the directories represent topics and each of them contains HTML files, translated in each language with a particular name.

The main advantage of this method is offered by the possibility to add a new translation without major changes on the web site infrastructure. However, there are many disadvantages that make this method not very attractive: difficult web site maintenance, possible differences between the site versions (in different languages); lose of the uniform image worldwide (for example a 10 web pages site will become a web site with 70 pages when it is translated in 6 languages). Any update or change must be done in each site in each language.

The *selective replication method* uses the same principle as the replication method. This method consists of partial replication of the web content. A pattern file is created for each web page. This file imports data from files with generic names and the

data is stored in a structure of directories based on the available languages. This method provides the preservation of the web site uniformity but still remains the site maintenance problem.

Dynamic Content Generation using Database is a method that can be implemented in certain conditions: (a) the web site host server must be fast, (b) the site shouldn't be too large and it will be translated only in a few languages (usually 2-3), (c) the graphical elements are few. The data is stored in tables; every message is identified by a unique ID.

The disadvantages of this method are presented in Table 1.

The last method, *Dynamic Content Generation using XML Files*, combines elements of the previous presented methods. In the frame of this method, a template is designed for each web page. This template generates the format for the web page displayed to the client. The texts and images for each language are maintained in a separate XML file. The XML files are stored in a structure of subdirectories according to the languages associated. The template file imports messages from XML files and generates the content displayed on the client.

Thus, this method is similar, in principle, with the previous method; the main difference is that the data is stored in XML files, not in a database. Storing data in XML files offers many advantages (see Table 1). This method requires a good initial strategy and organization plan. If this strategy is optimal developed, the site maintenance will not cause problems and adding a new language will not change in a great extent the site infrastructure.

2.1 Hybrid Method Architecture

The hybrid method architecture emerges elements from all the methods presented before. This new method proposes to valorise the advantages of the methods presented above and reduce their disadvantages. Table 1 explains how the disadvantages of one method can be eliminated, using the powerful characteristics of another one.

The disadvantages of the hybrid method can be eliminated by creating several web instruments for information uploading on the server and the web site maintain. The programming language selected to create the script and the server performances are also very important.

According to this method, a template file will be created for each web page. The template can be created using any programming language for CGI interfaces.

The template file will generate the web page format using the data extracted from the database and XML files.

Table 1: Hybrid method versus other methods.

Method	Advantages	Disadvantages
Replication (Site / Selective)	Increased number of languages; High performance - low server overhead.	The uniformity of the site is not preserved; Large number of web files to maintain.
Dynamic Content Generation using Database	Maintain the site uniformity using a web template; It is a good choice for small sites with reduced multimedia content.	Reduced number of languages; Slow performance on server; Difficult to load information in the database.
Dynamic Content Generation using XML Files	Better performance on server unlike dynamic content using database; It is a good choice for larger sites.	Can be implemented only at the initial phase of the design. Difficult to store information on site.
Hybrid method	There is no limit for numbers of languages; Better performance on server (use files on disk and database); The uniformity of the site is preserved; The web site is easy to maintain.	It is difficult to maintain the information on site without any additional instruments; Medium server overhead.

All the messages to be translated are stored in XML files and the database contains only general information used by the template script. In addition, the XML files are stored using a subdirectories-based architecture, in order to be easily accessed. The method architecture is presented in Figure 1.
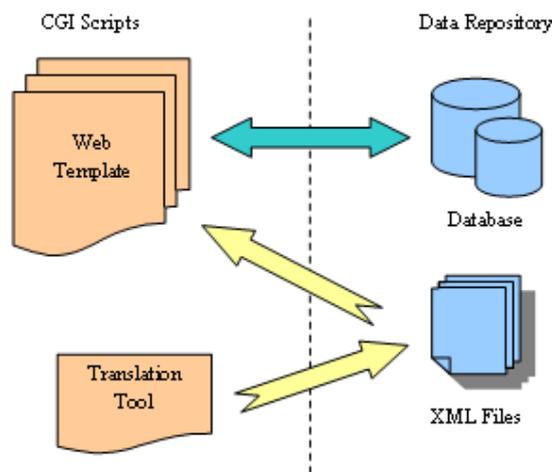

Figure 1: Hybrid method architecture.

The XML files are created using a web instrument - Translation tool. This instrument cannot

make the translation itself; it is used only by the translators to on-line introduce the translated web page content. Using this procedure, the web master doesn't interact with the translation web page content. In this case, the role of the persons involved in translation activities and site maintenance is clearly defined.

2.2 VccSSe Web Site Architecture

VccSSe (Virtual Community Collaborating Space for Science Education) - http://vccsse.ssai.valahia.ro - is a Comenius 2.1. Project funded by the European Commission. The project is aimed to adapt, develop, test, implement and disseminate training modules, teaching methodologies and pedagogical strategies based on the use of Virtual Instruments, with the view to their implementation in the classroom, through Information and Communication Technology tools. In this sense, the partnership assumes to build various pedagogical approaches in a virtual space able to offer efficient ways of using specific tools for logical understanding of the fundamental concepts in sciences. The target groups are formed by approximately 180 in-service teachers from primary and secondary schools involved in Sciences teaching areas in the partner countries (Romania, Spain, Poland, Finland and Greece), institutions' local coordinators, tutors, researchers, local educational authorities and over 3500 pupils who become indirect beneficiaries. In addition, teachers and trainers from Europe will benefit from the project outputs consisting of a number of on-line simulating laboratories.

Having in view the huge amount of transnational work and outputs, the main channel of dissemination is represented by the project web page. In this context, the whole on-line content - first created in English - had to be translated in other 5 languages: Romanian, Spanish, Polish, Finnish and Greek. The increased number of translation languages eliminates the possibility to implement the Dynamic Content Generation using Database method.

Using the hybrid method architecture described before, the translation process is simplified. The web masters developed for each web page a template using a powerful programming language: Perl. The web template is used only to generate the web page format and, in general, it doesn't contain any text that can be displayed in the web page. The web template queries a MySQL database to obtain any information needed to generate the web page or useful for other operations. The content stored in the database is not intended to be translated. All the

messages that will be displayed in the web page are stored by in a XML file, created for the English version of the web site. The XML files are stored in the subdirectories architecture based on languages acronym.

When a user accesses the web page, the web template is executed on the server. The Perl script contains a subroutine (presented below) that verifies the existence of the XML files for each language and creates a list with these languages.

```perl
sub langs {
  my $tran;
  my @flags =
("en","ro","es","pl","fi","gr");
  my $altlg =
{en=>"English",ro=>"Rom&#226;n&#259;",e
s=>"Español",pl=>"Polski",fi=>"Suomi",
gr=>"&#917;&#955;&#955;&#951;&#957;&#95
3;&#954;&#940;"};
  foreach (@flags) {
    if (-e
"../langs/".$_."/".basename($0).".xml")
  {$tran = $tran." ".$q->a({-
href=>"/main/".basename($0)."?lang=$_"}
, $q->img({-src=>"/images/$_.gif",-
border=>"0",-alt=>"$altlg->{$_}"})); }}
    return $q->div({-
id=>"langs"},"$mmsg->{al}$tran");}
```

Based on this list, an image of the flag and a link to the web page in that language is displayed. If the XML file, for a specific language, is missing, the associated flag is not displayed and the web page (in that language) cannot be accessed.

2.3 Results

To translate a web page from English in other language, the person in charge, must access the translation tool, available on the web page. For example, if a Romanian translator wants to translate the Introduction web page, he must access the Translation tool button, located on the English Introduction web page. After an authentication process, he is forced by the translation interface to translate only for the Romanian version of the page (this is accomplished by the authentication information stored in the MySQL database). The translation tool dynamically generates an interface based on the English messages stored in the XML file; the fields dimension where the Romanian messages will be written are proportionally calculated by the numbers of characters of the

English messages. These dimensions are only visual and don't restrict the translated messages size.

The translation tool has two purposes: to create the XML files that contain messages in specified language when it is used for the first time on a web page and to update messages on specified web page. The translators use the keyboard in their languages and the specials characters are dynamically converted by the CGI script in web codes; as a result of this feature the problem of the character set is eliminated – it is not necessary to generate web pages with a specific character set declaration.

The translation tool is the only instrument used by the partners to translate and update the web content in their languages. This Perl script (partially presented below) is very small - 4.1 Kbytes and contains only 345 words, being executed very fast by the server (a Dual Core CPU server).

```perl
 print $q->b("Language: $lang");
  print           $q->start_form(-name=>
"translate",-method=>"POST"),$q-
>start_table;
 my $count = 0, $bgcolor = '';
 foreach (keys %{$en}) {
  $count++;
  if ( $count/2 == int($count/2) ) {
$bgcolor = "bgcolor='#CCCCCC'"; } else
{ $bgcolor = ""; }
  my $lines = (length $en->{$_})/50;
    print           $q->Tr({$bgcolor},[$q-
>td(["$en->{$_}",
   $q->textarea(-name=>"$_",-
rows=>"$lines",-columns=>"50",-
value=>"$olang->{$_}")])]);}
   print     $q->end_table,$q->hidden(-
name=>'op',-default=>'s'),
    $q->hidden(-name=>'page',-
default=>$page),
    $q->submit(-name=>"",-
value=>"Save"),$q->end_form;
   print           $q->start_form(-
name=>"authen",-method=>"POST"),
    $q->hidden(-name=>'page',-
default=>$page),
    $q->hidden(-name=>'logout',-
default=>'yes'),
    $q->submit(-name=>"authen",-
value=>"Logout"),$q->end_form;
```

Using this hybrid method, the web site uniformity is preserved. The persons that translate the web content are forced to translate phrase by phrase; the text is presented in a random list. The number of people involved in this action is reduced and the possibility to appear errors is insignificant.

3 CONCLUSIONS

The hybrid method architecture for designing multi-language web sites was implemented in the Comenius 2.1 Project VccSSe (Virtual Community Collaborating Space for Science Education). The website is available since the beginning of the year 2007 and it can be accessed at: http://vccsse.ssai.valahia.ro. The web site is under development until the end of the project. It is easy to add new web pages or new languages for translation without making changes in the site architecture.

The English content posted on the site has been translated in all five languages of the project partners (Romanian, Spanish, Polish, Finnish and Greek) using only the translation tool. The translation tool – a very small and fast script - can be used to create new web pages in other languages or to update the online-content for a specific language. This tool cannot be used to change the English content. To fulfil this action, another web tool is developed.

ACKNOWLEDGEMENTS

REFERENCES

Itzkowitz, Z., Grunwald, D., link verified in 2008. *Considerations in the Design of Multi-Language Websites,* http://www.global-translation-services.com/download/Design_of_Multi-Language_Websites.pdf

Stylusinc, 2005, *How can your website speak multiple languages to your customers?*, StylusInc Pvt. Ltd., http://www.stylusinc.com/Common/whitepapers/WhitePapers/Multi-Language%20Support.pdf

Tiscali, 2008. *Tiscali Encyclopaedia*: http://www.tiscali.co.uk/reference/encyclopaedia/

VccSSe, 2008. *Virtual Community Collaborating Space for Science Education* – VccSSe European Project web site, http://www.vccsse.ssai.valahia.ro