

一种多语言网站解决方案^①

王 业¹, 杨 琴^{2,3}, 王亮亮¹

¹(新疆农业大学 现代教育技术中心, 乌鲁木齐 830052)

²(新疆教育学院 信息科学与技术学院, 乌鲁木齐 830043)

³(新疆教育学院 远程教育技术重点实验室, 乌鲁木齐 830043)

摘 要: 为了满足企业信息化和全球化的需要, 各大企业利用互联网技术为全球用户提供更为完善的服务. 提出一种基于资源文件的多语言网站的设计思想和总体架构, 采用 LAMP 技术, 设计并实现了多语言网站. 旨在为解决用户与企业交流问题提供了一个强有力的平台, 同时能够有效加速企业的全球化和本地化的发展进程, 为企业在全球贸易中获取更多的经济利益.

关键词: 多语言网站; LAMP

A Multi-Language Website Solution

WANG Ye¹, YANG Qin^{2,3}, WANG Liang-Liang¹

¹(Modern Education Technology Center, Xinjiang Agricultural University, Urumqi 830052, China)

²(College of Information Science and Technology, Xinjiang Education Institute, Urumqi 830043, China)

³(Key Laboratory of Distance Education Technology, Xinjiang Education Institute, Urumqi 830043, China)

Abstract: In order to meet the needs of information technology and globalization, the major companies use Internet technology to provide users with better service. This paper presents design principles and overall architecture of a multi-language website based on resource file. With WAMP technology, we design and implement a multi-language website, which designed to provide a strong platform to solve the problem of users and business exchanges. At the same time, this paper aimed to accelerate the development of globalization and localization processes effectively for enterprises in global trade to obtain more economic benefits.

Key words: a multi-language website solution; LAMP

1 引言

随着国际化贸易的飞速发展, 各大企业利用互联网技术为全球用户提供更为完善的服务. 在我国, 有很多地区是少数民族聚居地区, 存在不同民族的多语言环境. 面对来自各地的访问用户, 语言差异成为用户与企业之间交流的最大障碍. 多语言 Web 网站为解决交流问题提供了一个强有力的平台, 同时能够有效加速企业的全球化和本地化的发展进程, 为企业在全球贸易中获取更多的经济利益^[1].

本文针对新疆多语言文字的特点及民族地区对信息技术的需求现状, 在 Unicode 字符集的基础上探究

多语言兼容技术, 开发了支持维、哈、汉多语言信息网站. 从技术层面来看, 多语言网站不是简单的对汉语网站的翻译, 而是在充分分析维吾尔文和哈萨克文的基础上, 设计开发整个网站, 使新疆各少数民族得到更丰富、更优质的信息技术服务.

1 多语言网站分析

1.1 现状分析

目前, 网站多语言版化大致分为两类:

(1) 一般中小规模的商业性网站, 采用第三方服务, 利用现有的语言翻译程序, 自动完成整个页面文本及

① 基金项目:新疆维吾尔自治区自然科学基金(2013211B28)

收稿时间:2013-11-25;收到修改稿时间:2014-01-03

内容的翻译。例如, Google 的网站翻译器, 提供 60 多种语言版本, 设计者只需在网站上插入 Google 的脚本, 即可实现多语言翻译。

优点: 此类网站成本小, 见效快, 基本不用日后人工维护。由于程序翻译有 self learning 的功能, 所以从长远发展来看程序翻译的文章越多则翻译结果越准确。

缺点: 翻译精度不高, 翻译结果相对不可控, 对于网站的 CSS 技术要求比较高。

(2)对原有语言版本的镜像拷贝。这种镜像拷贝是基于网站整体结构上和数据上的, 例如, 网站本身是中文的, 现在需要制作其他语言版本, 那么就需要以中文版本的网站结构和数据为基础, 翻译出不同语言版本, 并最终确保所有 URL 链接的有效性。将不同语言的页面存放在不同的目录下, 各语言版本表现出的页面内容和布局并不完全一致。同时, 翻译工作作为人为处理。

优点: 操作简单, 语言准确, 显示直观。

缺点: 代码的复用性较低, 资源冗余, 日常维护更新比较繁琐。

结合两种分类情况分析, 在现有的条件下, 由于机器自动翻译的效果并不是太理想, 所以采用人工事前翻译的办法解决翻译问题。文中在分类 2 设计思想的基础上, 构造通用的基础模板页面, 针对各语言的差异, 形成不同的样式文件, 并根据用户的语种选择, 整合基础模板页面及样式文件, 实现页面整体布局的统一, 从而增强代码复用性, 减少资源冗余。

1.2 维哈语言特点

维吾尔文和哈萨克文属于阿尔泰语系, 其文字都借用了阿拉伯语和部分波斯文字母。维吾尔文共有 32 个字母, 哈萨克文共 33 个字母^[2]。由于维吾尔和哈萨克文字是一种手写体的文字, 根据在单词中的位置不同, 每个字母有单独形式、词首形式、词中形式、词尾形式等 4 种表现形式。书写时由该字符在单词中的位置决定显现形式。因此, 维吾尔文和哈萨克文字符在输入编辑时具有一些特殊性^[3], 具体表现为: (1)书写方向为从右到左, 行向为从上而下, 输入时光标移动方向与汉、英文书写方向相反, 这使维吾尔文、哈萨克文与汉、英文混合编辑时技术处理比较复杂; (2)维吾尔文的标点符号, 例如逗号、问号等与汉、英文符号方向相反^[4]。

1.3 字符集的影响

国内网站一般采用的是简体中文(GB2312)字符集, 而对于维吾尔和哈萨克语网站来说, 汉语网站的字符集不支持其语言, 使用中文字符集会出现无法兼容从而导致乱码的情况。所以对要提供维哈汉多语言版本的网站来说, 应该选择一种对汉文、维吾尔文和哈萨克文均支持的字符集^[5]。Unicode 是一种双字节编码机制的字符集, 不管是东方文字还是西方文字, 在 Unicode 中一律用两个字节来表示, 因而至少可以定义 65536 个不同的字符, 几乎可以涵盖目前世界上所有通用语言的每一种字符^[6]。这种通用的书写字符和文本的编码方案, 提供了操作多语言文本间交换的统一方法。所以就多语言版本的网站而言, Unicode 字符集是比较理想的选择。所以就多语言版本的网站而言, Unicode 字符集是比较理想的选择。

2 系统设计

2.1 设计思路

(1)模板图样制作。将图样中的图像与文本分离, 在多语言切换时, 不改变图像, 将对应语言与图像合成。同时, 考虑语言文字的长度, 预留足够的图片空间。

(2)模板页面制作。对图样进行切图, 将公用图片放置统一文件目录中, 供模板页面调用。

(3)样式制作。针对不同语言, 考虑语言差异(汉语书写: 由左至右, 维、哈语书写: 由右至左), 制作相应样式文件, 并建立相应语言的样式目录。

(4)数据库设计。设置对应语言字段。在同一个表中为每种语言建立独立的字段, 不同语言请求访问不同字段。

(5)脚本设计。将各个模块的通用信息存放在统一的文件中, 便于代码复用、全局调用及管理; 同时, 在控制脚本中, 使用变量来代替字符串。

2.2 框架设计

采用 LAMP(Linux+Apache+MySQL+PHP)开发平台, 搭建一个三层架构系统, 下至上分别为数据访问层、业务逻辑层、表示层。

数据访问层。通过操纵数据为事务逻辑层提供数据服务, 例如存储数据、操作结果、返回数据检索结果等。

业务逻辑层。是整个系统的核心, 负责接收浏览

器传来的请求,并将请求传给数据层,同时将请求处理结果发给浏览器,由表单组成.表单是应用程序的核心所在,是向客户呈现数据和信息的基础,也是响应及处理客户与表单交互生成信息和数据的基础.

表示层.是系统的 UI 部分,负责用户与整个系统的交互,接收用户的输入并将服务器端传来的数据呈现给用户.获取客户端在进行 HTTP 请求时发送的信息,得到客户端使用的区域语言,并根据所用的区域语言自动读取相应的资源文件,进而生成不同语言的页面^[6].

2.3 页面请求过程

页面请求的过程可描述如图 1 所示:

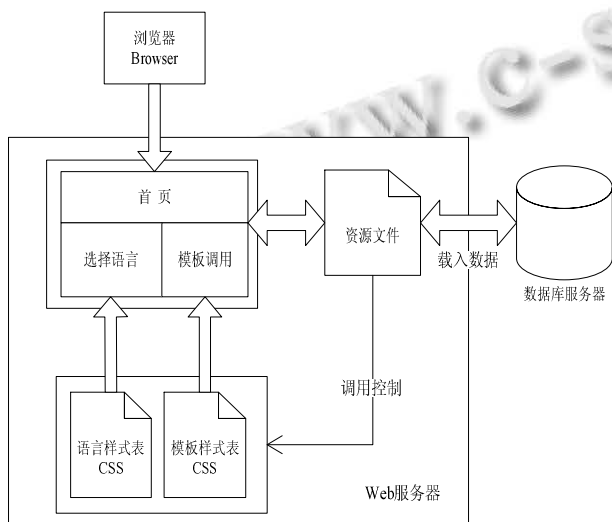


图 1 页面请求过程图

(1)用户通过浏览器访问站点,选择语言版本,并通过浏览器向服务器提交页面请求.

(2)Web 服务器响应用户的语言请求,由控制脚本从资源文件中加载相应语言的静态文本及 CSS 样式文件,进行页面初始化.

(3)控制脚本设置其对应的语言字段,作为参数传入数据库,检索对应语言的数据,进行动态数据提取.

(4)模板页面将接收到的各种信息,构成 html 并返回客户端浏览器,进行展示.

2.4 数据库设计

数据库采用字段级别来支持多语言,即在同一个表中为每种语言建立独立的字段,不同语言的用户操作不同的字段^[7].保留各种语言通用的字段,对于不通用的字段,分别建立不同语言的相关字段.数据库

的字符集采用 UTF-8.

表 1 频道信息表

字段	字符集	说明
CHANNEL_ID	utf8_general_ci	栏目 ID
CHANNEL_NAME	utf8_general_ci	栏目名称(中文)
uyg_CHANNEL_NAME	utf8_general_ci	栏目名称(维文)
kaz_CHANNEL_NAME	utf8_general_ci	栏目名称(哈文)
CHANNELTYPE	utf8_general_ci	栏目类型

3 主要实现方法

3.1 抽取语言文本

将页面中使用的静态文本提取出来,根据不同的语言,以数组的形式,分别放置在不同语言文件中,中文为 zh_CN.php,哈文为 kk_CN.php,维文为 ug_CN.php.其中,关于首页页面的语言文件如下:

(1)中文语言文件: zh_CN.php.部分内容如下:

//栏目

```
$DBTable['CHANNEL_NAME']='CHANNEL_NAME';
```

//首页静态文本

```
$Html['index'][1]='资源搜索';
$Html['index'][2]='其他栏目';
$Html['index'][3]='课件频道';
$Html['index'][4]='资源共享';
$Html['index'][5]='当前位置';
$Html['index'][6]='资源浏览';
$Html['index'][7]='节目单';
```

.....

(2)维文语言文件: ug_CN.php.部分内容如下:

//栏目

```
$DBTable['CHANNEL_NAME']='uyg_CHANNEL_NAME';
```

//首页静态文本

```
$Html['index'][1]='مەنبە ئىزدەش';
$Html['index'][2]='باشقا سەھىپىلەر';
$Html['index'][3]='دەرسلىك قانلى';
$Html['index'][4]='بايلىق كۆرگەنمىسى';
$Html['index'][5]='ھازىرقى ئورنىڭىز';
$Html['index'][6]='كۆز يۆگۈرۈش';
$Html['index'][7]='پىروگرامما تىزىملىكى';
```

.....

(3)哈文语言文件: kk_CN.php.部分内容如下:

//栏目

```
$DBTable['CHANNEL_NAME']='kaz_CHANNEL_
NAME';
```

```
//首页静态文本
```

```
$Html['index'][1]='ماتەريال قاينارەن مۇدەت';
```

```
$Html['index'][2]='باسقا تاقىرىپ';
```

```
$Html['index'][3]='ساباقتىق قۇرال ارناسى';
```

```
$Html['index'][4]='بايلىق قاينارەن تاننىق';
```

```
$Html['index'][5]='ازىرگى ورنى';
```

```
$Html['index'][6]='قاينار شولۇ';
```

```
$Html['index'][7]='باھادار لامالار';
```

```
.....
```

3.2 语言切换

用户选择语言时,根据提交语言请求,动态加载语言文件,并重新刷新页面,显示对应语言信息.同时,在 session 保存所选语言信息,以保证在页面跳转时获取对应语言文本.切换流程如下:

(1)获取用户选择语言请求,并在 session 中保存语言请求;

(2)在模板页面基础上,调用语言资源文件;

(3)通过脚本控制,动态整合对应的语言样式文件,构成页面结构主体.

(4)传递语言参数,通过数据库接口,提取动态数据进行内容填充.

3.3 数据提取

除了抽取页面显示的静态文本之外,还需要提取数据库中存储的多语言信息.将信息提取过程写入函数,存储在统一函数库文件中,在页面中进行模板置标调用.例如,“栏目展示”代码如下:

```
/*=====
```

```
管理页面-栏目(类别)菜单调用
```

```
$Lang: 语言种类,要调用栏目属于第几种语言
```

```
=====*/
```

```
function GetColumn_model($channelid){
```

```
global $db,$Lang,$DBTable;
```

```
$Rs=GetColumn_array($channelid);
```

```
$channelname=$DBTable['CHANNEL_NAME'];
```

```
// lib/*_CN.php 语言文件中获取字段名称
```

```
for ($i=0;$i<count($Rs);$i++) {
```

```
    $CHANNEL_ID = $Rs[$i][CHANNEL_ID];
```

```
    $CHANNEL_NAME =
```

```
$Rs[$i][$channelname];
```

```
    echo '<div class="two">';
```

```
    echo '<div class="top"><a
```

```
title="',$CHANNEL_NAME.'">
```

```
href="othernewsindex.php?channelid='.$CHANNEL_ID.
```

```
'>',$CHANNEL_NAME.'</a></div>';
```

```
    echo '<div class="end" >';
```

```
    GetColumnByID($CHANNEL_ID,"othernewslist.p
hp");
```

```
    echo '</div>';
```

```
    echo '</div>';
```

```
}
```

```
unset($Rs,$i,$CHANNEL_ID,$CHANNEL_NAME,
E,$channelname);
```

```
}
```

3.4 页面数据展示

(1)静态文字调用

展示页面中,根据用户选择的语言信息,调用语言文件中对应的静态文字内容.部分调用脚本如下:

```
<div class="end" id="div1" >
```

```
<ul>
```

```
<li><a
```

```
href="othernewsindex.php?channelid=A4012">
```

```
<?php echo $Html['index'][17]; ?></a>
```

```
</li>
```

```
<li><a href="othernewsindex.php?channelid=A5">
```

```
<?php echo $Html['index'][19]; ?></a>
```

```
</li>
```

```
<li><a href="othernewsindex.php?channelid=BY">
```

```
<?php echo $Html['index'][20]; ?></a>
```

```
</li>
```

```
</ul>
```

```
</div>
```

(2)动态数据调用

除了静态文字调用之外,还有读取数据库中的动态内容.展示页面中,根据用户选择的语言信息,调用读取动态信息的公用置标函数.部分调用脚本如下:

```
<table width="100%" height="280" border="0"
cellpadding="0" cellspacing="0" bgcolor="#EEF2FB">
```

```
<tr>
```

```
<td width="282" valign="top">
```

```
<div id="container">
```

```
<?=GetColumn_man()?>
```

```
</div>
</td>
</tr>
</table>
```

4 结语

文中主要讨论了多语言 Web 网站的设计原则和思路,并结合实例,详细地描述了多语言网站的结构和实现方案.该方案解决了以往网站框架和内容相互依赖,网站本地化过程复杂的问题.资源文件的使用进一步简化了网站内容的管理,站点维护人员无需了解网站的内部实现,即可实施站点的及时更新,在一定程度上提高了网站开发的效率,降低了后期维护的成本,为多语言网站解决方案信息化管理的实现提供了良好的前景.

参考文献

- 1 张菲菲,薛贺,李建良.多语言 Web 网站的设计与实现.微电子学与计算机,2008,(5):43-45.
- 2 买买提艾力,佟加·庆夫,亚森·伊明.信息交换用维吾尔文、哈萨克文、柯尔克孜文字体字形标准研究.语言与翻译(汉文),2005(4):51-53.
- 3 易坤诱,高士杰.维吾尔语语法.北京:中央民族大学出版社,1998.
- 4 马树才,黄海龙,蒋平安,张霞,李永.维、哈、汉多语言农业信息网站的设计及开发.广西农业科学,2009,(8):1101-1104.
- 5 瓦热斯江·阿布都克力木.维文 Unicode 在线处理技术的实现[学位论文].乌鲁木齐:新疆大学,2002.
- 6 张亚娟,付辉,王学春.基于动态技术的多语言网站研究与设计.现代计算机,2010,(3):128-130.
- 7 李谨.浅析基于 JSP 的维、汉双语网站的开发技术.科技创新导报,2009,(2):32.