使用PHP实现汉字转拼音

# 开发环境

1、环境搭建：Windows 7+Apache 2.4.18+MySQL 5.7.11+PHP 7.1.0。

2、文本编辑器：Sublime 3。

# 二、主要技术

本案例主要在PHP中基于GB2312类库，通过拼音搜索的实现原理从而来实现汉字转拼音的功能。

# 三、效果图展示

实现效果图，如图1所示。

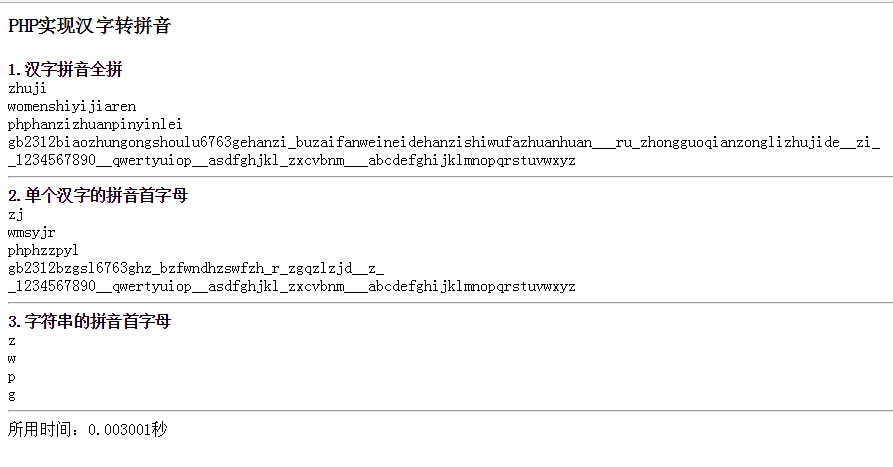


图1 效果图（汉字转拼音转换结果）

# 四、具体步骤

## 4.1、创建PHP文件

创建一个名为“index.php”的PHP文件。

## 4.2、汉字转拼音功能实现

### （1）创建汉字转拼音类

创建一个名为“pinyin”的类，创建代码如下：

class pinyin{

...

}

### （2）将utf8编码转换为GB2312

public static function utf8\_to\_gb2312($s) {

return iconv('UTF-8', 'GB2312//IGNORE', $s);

}

### （3）获得拼音数组

public static function get\_pinyin\_array() {

static $py\_arr;

if(isset($py\_arr)) return $py\_arr;

$k = 'a|ai|an|ang|ao|ba|bai|ban|bang|bao|bei|ben|beng|

bi|bian|biao|bie|bin|bing|bo|bu|ca|cai|can|cang|cao|ce|

ceng|cha|chai|chan|chang|chao|che|chen|cheng|chi|chong|

chou|chu|chuai|chuan|chuang|chui|chun|chuo|ci|cong|cou|

cu|cuan|cui|cun|cuo|da|dai|dan|dang|dao|de|deng|di|dian

|diao|die|ding|diu|dong|dou|du|duan|dui|dun|duo|e|en|er

|fa|fan|fang|fei|fen|feng|fo|fou|fu|ga|gai|gan|gang|gao

|ge|gei|gen|geng|gong|gou|gu|gua|guai|guan|guang|gui|gu

n|guo|ha|hai|han|hang|hao|he|hei|hen|heng|hong|hou|hu|h

ua|huai|huan|huang|hui|hun|huo|ji|jia|jian|jiang|jiao|j

ie|jin|jing|jiong|jiu|ju|juan|jue|jun|ka|kai|kan|kang|k

ao|ke|ken|keng|kong|kou|ku|kua|kuai|kuan|kuang|kui|kun|

kuo|la|lai|lan|lang|lao|le|lei|leng|li|lia|lian|liang|l

iao|lie|lin|ling|liu|long|lou|lu|lv|luan|lue|lun|luo|ma

|mai|man|mang|mao|me|mei|men|meng|mi|mian|miao|mie|min|

ming|miu|mo|mou|mu|na|nai|nan|nang|nao|ne|nei|nen|neng|

ni|nian|niang|niao|nie|nin|ning|niu|nong|nu|nv|nuan|nue

|nuo|o|ou|pa|pai|pan|pang|pao|pei|pen|peng|pi|pian|piao

|pie|pin|ping|po|pu|qi|qia|qian|qiang|qiao|qie|qin|qing

|qiong|qiu|qu|quan|que|qun|ran|rang|rao|re|ren|reng|ri|

rong|rou|ru|ruan|rui|run|ruo|sa|sai|san|sang|sao|se|sen

|seng|sha|shai|shan|shang|shao|she|shen|sheng|shi|shou|

shu|shua|shuai|shuan|shuang|shui|shun|shuo|si|song|sou|

su|suan|sui|sun|suo|ta|tai|tan|tang|tao|te|teng|ti|tian

|tiao|tie|ting|tong|tou|tu|tuan|tui|tun|tuo|wa|wai|wan|

wang|wei|wen|weng|wo|wu|xi|xia|xian|xiang|xiao|xie|xin|

xing|xiong|xiu|xu|xuan|xue|xun|ya|yan|yang|yao|ye|yi|yi

n|ying|yo|yong|you|yu|yuan|yue|yun|za|zai|zan|zang|zao|

ze|zei|zen|zeng|zha|zhai|zhan|zhang|zhao|zhe|zhen|zheng

|zhi|zhong|zhou|zhu|zhua|zhuai|zhuan|zhuang|zhui|zhun|z

huo|zi|zong|zou|zu|zuan|zui|zun|zuo';

$v = '-20319|-20317|-20304|-20295|-20292|-20283|-20265

|-20257|-20242|-20230|-20051|-20036|-20032|-20026|-

20002|-19990|-19986|-19982|-19976|-19805|-19784|-19775

|-19774|-19763|-19756|-19751|-19746|-19741|-19739|-

19728|-19725|-19715|-19540|-19531|-19525|-19515|-19500

|-19484|-19479|-19467|-19289|-19288|-19281|-19275|-

19270|-19263|-19261|-19249|-19243|-19242|-19238|-19235

|-19227|-19224|-19218|-19212|-19038|-19023|-19018|-

19006|-19003|-18996|-18977|-18961|-18952|-18783|-18774

|-18773|-18763|-18756|-18741|-18735|-18731|-18722|-

18710|-18697|-18696|-18526|-18518|-18501|-18490|-18478

|-18463|-18448|-18447|-18446|-18239|-18237|-18231|-

18220|-18211|-18201|-18184|-18183|-18181|-18012|-17997

|-17988|-17970|-17964|-17961|-17950|-17947|-17931|-

17928|-17922|-17759|-17752|-17733|-17730|-17721|-17703

|-17701|-17697|-17692|-17683|-17676|-17496|-17487|-

17482|-17468|-17454|-17433|-17427|-17417|-17202|-17185

|-16983|-16970|-16942|-16915|-16733|-16708|-16706|-

16689|-16664|-16657|-16647|-16474|-16470|-16465|-16459

|-16452|-16448|-16433|-16429|-16427|-16423|-16419|-

16412|-16407|-16403|-16401|-16393|-16220|-16216|-16212

|-16205|-16202|-16187|-16180|-16171|-16169|-16158|-

16155|-15959|-15958|-15944|-15933|-15920|-15915|-15903

|-15889|-15878|-15707|-15701|-15681|-15667|-15661|-

15659|-15652|-15640|-15631|-15625|-15454|-15448|-15436

|-15435|-15419|-15416|-15408|-15394|-15385|-15377|-

15375|-15369|-15363|-15362|-15183|-15180|-15165|-15158

|-15153|-15150|-15149|-15144|-15143|-15141|-15140|-

15139|-15128|-15121|-15119|-15117|-15110|-15109|-14941

|-14937|-14933|-14930|-14929|-14928|-14926|-14922|-

14921|-14914|-14908|-14902|-14894|-14889|-14882|-14873

|-14871|-14857|-14678|-14674|-14670|-14668|-14663|-

14654|-14645|-14630|-14594|-14429|-14407|-14399|-14384

|-14379|-14368|-14355|-14353|-14345|-14170|-14159|-

14151|-14149|-14145|-14140|-14137|-14135|-14125|-14123

|-14122|-14112|-14109|-14099|-14097|-14094|-14092|-

14090|-14087|-14083|-13917|-13914|-13910|-13907|-13906

|-13905|-13896|-13894|-13878|-13870|-13859|-13847|-

13831|-13658|-13611|-13601|-13406|-13404|-13400|-13398

|-13395|-13391|-13387|-13383|-13367|-13359|-13356|-

13343|-13340|-13329|-13326|-13318|-13147|-13138|-13120

|-13107|-13096|-13095|-13091|-13076|-13068|-13063|-

13060|-12888|-12875|-12871|-12860|-12858|-12852|-12849

|-12838|-12831|-12829|-12812|-12802|-12607|-12597|-

12594|-12585|-12556|-12359|-12346|-12320|-12300|-12120

|-12099|-12089|-12074|-12067|-12058|-12039|-11867|-

11861|-11847|-11831|-11798|-11781|-11604|-11589|-11536

|-11358|-11340|-11339|-11324|-11303|-11097|-11077|-

11067|-11055|-11052|-11045|-11041|-11038|-11024|-11020

|-11019|-11018|-11014|-10838|-10832|-10815|-10800|-

10790|-10780|-10764|-10587|-10544|-10533|-10519|-10331

|-10329|-10328|-10322|-10315|-10309|-10307|-10296|-

10281|-10274|-10270|-10262|-10260|-10256|-10254';

$key = explode('|', $k);

$val = explode('|', $v);

$py\_arr = array\_combine($key, $val);

arsort($py\_arr);

return $py\_arr;

}

### （4）将汉字转换为拼音

// 汉字转换到拼音(字符串必须为GB2312编码)

public static function to($s, $isfirst = false) {

$res = '';

$len = strlen($s);

$pinyin\_arr = self::get\_pinyin\_array();

for($i=0; $i<$len; $i++) {

$ascii = ord($s[$i]);

if($ascii > 0x80) {

$ascii2 = ord($s[++$i]);

$ascii = $ascii \* 256 + $ascii2 - 65536;

}

if($ascii < 255 && $ascii > 0) {

if(($ascii >= 48 && $ascii <= 57) || ($ascii >= 97 && $ascii <= 122)) {

$res .= $s[$i]; // 0-9 a-z

}elseif($ascii >= 65 && $ascii <= 90) {

$res .= strtolower($s[$i]); // A-Z

}else{

$res .= '\_';

}

}elseif($ascii < -20319 || $ascii > -10247) {

$res .= '\_';

}else{

foreach($pinyin\_arr as $py=>$asc) {

if($asc <= $ascii) {

$res .= $isfirst ? $py[0] : $py;

break;

}

}

}

}

return $res;

}

### （5）输出汉字所对应的拼音

// 输出汉字所对应的拼音($isfirst代表输出汉字全拼或拼音首字母)

public static function utf8\_to($s, $isfirst = false) {

return self::to(self::utf8\_to\_gb2312($s), $isfirst);

}

### （6）输出字符串的拼音首字母

public static function to\_first($s) {

$ascii = ord($s[0]); // 返回字符串的首个字符的ASCII值

if($ascii > 0xE0) {

$s = self::utf8\_to\_gb2312($s[0].$s[1].$s[2]);

}elseif($ascii < 0x80) {

if($ascii >= 65 && $ascii <= 90) {

return strtolower($s[0]);

}elseif($ascii >= 97 && $ascii <= 122) {

return $s[0];

}else{

return false;

}

}

if(strlen($s) < 2) {

return false;

}

$asc = ord($s[0]) \* 256 + ord($s[1]) - 65536;

// 将ASCII值转换成拼音

if($asc>=-20319 && $asc<=-20284) return 'a';

if($asc>=-20283 && $asc<=-19776) return 'b';

if($asc>=-19775 && $asc<=-19219) return 'c';

if($asc>=-19218 && $asc<=-18711) return 'd';

if($asc>=-18710 && $asc<=-18527) return 'e';

if($asc>=-18526 && $asc<=-18240) return 'f';

if($asc>=-18239 && $asc<=-17923) return 'g';

if($asc>=-17922 && $asc<=-17418) return 'h';

if($asc>=-17417 && $asc<=-16475) return 'j';

if($asc>=-16474 && $asc<=-16213) return 'k';

if($asc>=-16212 && $asc<=-15641) return 'l';

if($asc>=-15640 && $asc<=-15166) return 'm';

if($asc>=-15165 && $asc<=-14923) return 'n';

if($asc>=-14922 && $asc<=-14915) return 'o';

if($asc>=-14914 && $asc<=-14631) return 'p';

if($asc>=-14630 && $asc<=-14150) return 'q';

if($asc>=-14149 && $asc<=-14091) return 'r';

if($asc>=-14090 && $asc<=-13319) return 's';

if($asc>=-13318 && $asc<=-12839) return 't';

if($asc>=-12838 && $asc<=-12557) return 'w';

if($asc>=-12556 && $asc<=-11848) return 'x';

if($asc>=-11847 && $asc<=-11056) return 'y';

if($asc>=-11055 && $asc<=-10247) return 'z';

return false;

}

## 4.3、汉字转拼音功能测试代码

### （1）输出汉字的拼音全拼

echo "<b>1.汉字拼音全拼</b><br/>";

print\_r(pinyin::utf8\_to('朱镕基'));

echo '<br/>';

print\_r(pinyin::utf8\_to('我们是一家人'));

echo '<br/>';

print\_r(pinyin::utf8\_to('PHP汉字转拼音类'));

echo '<br/>';

print\_r(pinyin::utf8\_to('GB2312标准共收录6763个汉字，不在范围内的汉字是无法转换，如：中国前总理朱镕基的“镕”字。'));

echo '<br/>';

print\_r(pinyin::utf8\_to('`1234567890-=QWERTYUIOP[]ASDFGHJKL;

ZXCVBNM,./abcdefghijklmnopqrstuvwxyz'));

echo '<hr/>';

### （2）输出单个汉字的拼音首字母

echo "<b>2.单个汉字的拼音首字母</b><br/>";

print\_r(pinyin::utf8\_to('朱镕基', 1));

echo '<br/>';

print\_r(pinyin::utf8\_to('我们是一家人', 1));

echo '<br/>';

print\_r(pinyin::utf8\_to('PHP汉字转拼音类', 1));

echo '<br/>';

print\_r(pinyin::utf8\_to('GB2312标准共收录6763个汉字，不在范围内的汉字是无法转换，如：中国前总理朱镕基的“镕”字。', 1));

echo '<br/>';

print\_r(pinyin::utf8\_to('`1234567890-=QWERTYUIOP[]ASDFGHJKL;

ZXCVBNM,./abcdefghijklmnopqrstuvwxyz', 1));

echo '<hr/>';

### （3）输出字符串的拼音首字母

echo "<b>3.字符串的拼音首字母</b><br/>";

print\_r(pinyin::to\_first('朱镕基'));

echo '<br/>';

print\_r(pinyin::to\_first('我们是一家人'));

echo '<br/>';

print\_r(pinyin::to\_first('PHP汉字转拼音类'));

echo '<br/>';

print\_r(pinyin::to\_first('GB2312标准共收录6763个汉字，不在范围内的汉字是无法转换，如：中国前总理朱镕基的“镕”字。'));

echo '<hr/>';

# 五、网页测试

开启Apache服务器后，打开浏览器，在浏览器地址栏中输入文件地址，例如：“localhost:8888/PHParticle/098/Codes/

index.php”，可以看到程序运行结果如图2所示。

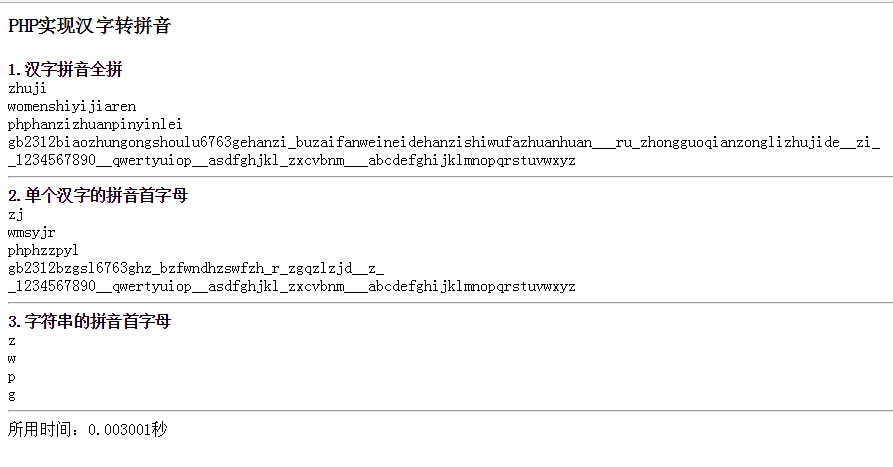


图2 案例结果图（汉字转拼音转换结果）