
utf8 support?

Posted by qqwwong - 2007/06/12 14:37

My Joomla contents are in utf8. I want to transfer them to another installation. Does the export content support utf8?

=====

Re: utf8 support?

Posted by Steve - 2007/06/12 15:36

When export content compiles the data it changes any non English characters to their ASCII numeric equivalent so that it won't be rejected by the install process.

I'm not sure how well it works on anything other than Latin based languages as I have not received any feedback.

If it doesn't find anything that is from the ASCII extended range it will parse it as is and should do so using your charset.

Please back up first though just to be safe.

Please let me know how you go.

I'm working on the next version for Joomla 1.5+ series and any feedback would be welcomed. :)

=====

Re: utf8 support?

Posted by jaleo - 2008/03/05 03:04

Hello,administraor.

My site content include chinese character.
It can not be exported correctly.

any adea ?

=====

Re: utf8 support?

Posted by montemedia - 2008/03/05 06:27

hello Steve, I guess I have problem with utf-8 encoding.

Joomla is 1.0.15, and com_export_content 1.7

I can install component and generate import_content_data.zip file.

Problem is that I can not install imported content as component at all.

Error is: can not find installation file.

I tried both with installing zip, or unpacking it into proper directory.
There is no problem with permissions, etc. I am rather sure that problem is in .xml file.

Would you be so kind to assist?

=====

Re: utf8 support?

Posted by Steve - 2008/03/05 17:48

jaleo
My site content include chinese character.

It can not be exported correctly.
What version of Export Content are you using.
What version are you compiling for 1.0+ or 1.5+?
And how are you compiling by sections or static content items?
If by sections are you including static and or archived items?
Do the "download" and "remove all" icons appear at the top after you press "compile"?

montemedia

Have you converted Joomla 10+series install over to UTF8?
If so export_content 1.7 assumes that it is not UTF8 and will try to convert all characters to their numeric equivalent however if it is already utf-8 encoded you will run into problems.

I'm guessing you are trying to move content over to another Joomla 1.0+ install
If this is the case try to move just one static item and see how it goes.
If it fails please post the file here from the trouble shooting section under the download button.
There have been some problems with attachments on the forums so if you just give me a link to the file or the compiled data component I will have a look at it.:)

Re:utf8 support?

Posted by montemedia - 2008/03/05 18:49

Hello Steve,

Thanks for prompt answer, and thanks for your efforts. I appreciate it. You probably can not remember, but you helped me a lot about year ago with similar problem -- exporting of UTF-8 content.
Now, it is time do do it again :)

Here is an update of my "situation": I can import static content, but can not content with sections/categories. Although exported content looks weird, after importing it into new site, UTF-8 looks perfect, and it is fine.

When I create import_content file from sections/categories content, I have an error "no xml file...".

Please, contact me on my e-mail: dragan68cg.yu, and I will provide you both working and non working file. Please understand that I have more than 1000 articles and sharing download link around sounds like not so good idea. Thanks for understanding.

Best,
Dragan

Re:utf8 support?

Posted by jaleo - 2008/03/05 21:12

Steve wrote:
jaleo
My site content include chinese character.

It can not be exported correctly.
What version of Export Content are you using.
What version are you compiling for 1.0+ or 1.5+?
And how are you compiling by sections or static content items?
If by sections are you including static and or archived items?
Do the "download" and "remove all" icons appear at the top after you press "compile"?

Hello, administrator.

Export Content 1.7, Joomla 1.0.13.
export content by sections and do not include static and or archived items, and to compiling to 1.5+
I can see "download " icons and can download it.
but after i import on joomla 1.5.1 site, chinese characters are not correct, like" & # 8217".

=====

Re:utf8 support?

Posted by Steve - 2008/03/05 22:12

Hi jaleo

Is it traditional or simplified.

If you can PM me a link to the chinese alphabet you are using and when I get time I will try and put a patch together.:)

=====

Re:utf8 support?

Posted by jaleo - 2008/03/05 22:51

Source site: <http://www.Bluelibrary.net> (Joomla 1.0.13)
Target site: <http://www.Bluelibrary.net/jm15/index> (Joomla 1.5.1 you can see simplified chinese character are not correct, such as "Godaddy æS-æ Úä¼Üæ ")

and what's your email? i want to send you exported data.

=====

Re:utf8 support?

Posted by Steve - 2008/03/06 16:17

Hi jaleo

When you look at content items under the edit window in Export Content are the characters corrupted or are they as they should be?

I had a look at the destination site and changed the encoding in the browser to traditional which does get most characters up although I don't know if they are correct as I can't read Chinese.

Have you tried another template just to make sure?

and what's your email? i want to send you exported data.
steve at bestdownloadsites.com put Export content in the subject line. :)

=====

Re:utf8 support?

Posted by jaleo - 2008/03/06 21:14

i cant read it after i choose traditional chinese encode,
i have send you email today. please check it.
Thanks.

=====

Re:utf8 support?

Posted by Steve - 2008/03/08 01:47

I have just had a look at the generated sql and can see what the problem is.

Export content assumes that content compiled on Joomla 1.0+ is not utf8 and will attempt to convert it over.

So on the source site if you go to:

/administrator/components/com_export_content/convert_img.php

This is where all the character conversions take place when you press the compile buttons.

You will see about line 23 function numeric_html(\$text){ this function goes down to about line 2208 you need to remove everything in between as shown below:

```
function numeric_html($text){  
REMOVE CODE  
return strtr($text, $cr);  
//return $text;  
}
```

And then just under that about line 2210 you will see:

```
/******  
High end numeric converter  
*****/  
function high_order($text){  
$text = preg_replace('/(./se', "'&#' . ((ord(substr('$1', 0, 1)) - 192) * 64 + (ord(substr('$1', 1, 1)) - 128)) . '";  
$text = preg_replace('/(..)/se', "'&#' . ((ord(substr('$1', 0, 1)) - 224) * 4096 + (ord(substr('$1', 1, 1)) - 128) * 64 +  
(ord(substr('$1', 2, 1)) - 128)) . '";  
return $text;  
}
```

```
function high_order($text){  
return $text;  
}
```

```
function high_order($text){  
return $text;  
}
```

```
function high_order($text){  
$text = preg_replace('/(./se', "'&#' . ((ord(substr('$1', 0, 1)) - 192) * 64 + (ord(substr('$1', 1, 1)) - 128)) . '";  
$text = preg_replace('/(..)/se', "'&#' . ((ord(substr('$1', 0, 1)) - 224) * 4096 + (ord(substr('$1', 1, 1)) - 128) * 64 +  
(ord(substr('$1', 2, 1)) - 128)) . '";  
return $text;  
}
```

```
return $text;  
}
```

Change this to:

```
function high_order($text){  
return $text;  
}
```

Once you change this it should compile all your Chinese characters as they are now in the source site. Please let me know how you go. :)

Re:utf8 support?

Posted by jaleo - 2008/03/08 09:54

Export empty content.

Re:utf8 support?

Posted by montemedia - 2008/03/08 13:13

Hello Steve,

The same situation is here. Now I can install exported content, but, actually, there is nothing to import - no content.

I have searched through my previously exported content and have noticed a lot of strange characters, with no special rule to repeat.

I have compared it with UTF-8 table I have, a lot of them does not have responding numeric code.

As far as my language is concerned, and the most of Middle EU countries, we need only 10-12 characters: ž, , , š, and responding capital letters.

Re:utf8 support?

Posted by Steve - 2008/03/09 19:14

Sorry it took so long to get back guys I have been very busy with paid work.

jaleo

Sorry the first one should look like this:

```
function numeric_html($text){  
REMOVE CODE  
return $text;  
}
```

And the second like this:

```
function high_order($text){  
return $text;  
}
```

That should bypass any character conversion.

montemedia

You will need to add a bit of code because when moving to Joomla 1.0+ series

we are using an XML install and XML will not parse certain characters so they must be converted.

So if you only want to move Serbian, Croatian, Montenegrin you need change the first function like so:

```
function numeric_html($text){  
REMOVE CODE  
return $text;  
}
```

If you paste it in make sure you don't leave the "REMOVE CODE" text there.

And the second one like this:

```
/******  
High end numeric converter  
*****/  
function high_order($text){  
$text = preg_replace('/(.)se', '&#x' . ((ord(substr('$1', 0, 1)) - 192) * 64 + (ord(substr('$1', 1, 1)) - 128)) . ';', $text);  
$text = preg_replace('/(.)se', '&#x' . ((ord(substr('$1', 0, 1)) - 224) * 4096 + (ord(substr('$1', 1, 1)) - 128) * 64 +  
(ord(substr('$1', 2, 1)) - 128)) . ';', $text);  
return $text;  
}
```

```
function high_order($text){
```

```
$text = preg_replace('/(.)se', '&#x' . ((ord(substr('$1', 0, 1)) - 192) * 64 + (ord(substr('$1', 1, 1)) - 128)) . ';', $text);  
$text = preg_replace('/(.)se', '&#x' . ((ord(substr('$1', 0, 1)) - 224) * 4096 + (ord(substr('$1', 1, 1)) - 128) * 64 +  
(ord(substr('$1', 2, 1)) - 128)) . ';', $text);  
return $text;  
}
```

You should just copy-paste the attached code and replace the above function in convert_img.php.

http://www.bestdownloadsites.com/export_content/images/fbfiles/files/Serbian-Croatian-Montenegrin.txt

Re:utf8 support?

Posted by montemedia - 2008/03/09 19:38

Excuse me, Steve... where is the "Attached code"?

EDIT: URL with code suddenly appeared :) Thanks, I will try and inform you.

EDIT 2: Sorry Steve, again - nothing to import.

How to say... XML generates exact number of articles, but nothing inside - neither title, nor content.

I edited your txt file, adding some " " characters in front of Central European letters.

BTW, this is for Slovenian and Bosnian language, too. I am not sure about Slovak and Czech language, but I will check it up and inform you.

Re:utf8 support?

Posted by montemedia - 2008/03/09 20:27

Both Czech and Slovak are a little bit different regarding letter than Serbian, Slovenian, Croatian, Bosnian and Montenegrin.

I played a bit with your codes and tips: I added letter replacements in first edited part, and I have got both content and titles.

But, encoding is not good.

We have to put "something" between lines 23 and 24 :)

```
function numeric_html($text){
```

```
    PUT SOMETHING HERE, BUT NOT THIS:
```

```
    //start of not this
```

```
    $scr=array(
```

```
    //Serbian, Croatian, Montenegrin, Bosnian, Slovenian
```

```
    ' ' => ' ',
```

```
    ' ' => ' ',
```

```
    ' ' => ' ',
```

```
    ' ' => ' ',
```

```
    ' ' => ' ',
```

```
    ' ' => ' ',
```

```
    'Š' => 's',
```

```
    'š' => 'a',
```

```
    'Ž' => 'j',
```

```
    'ž' => 'r'
```

```
    );
```

```
    $text= strtr($text, $scr);
```

```
    return $text;
```

```
}
```

```
    // end of not this
```

```
    return $text;
```

```
}
```

Re:utf8 support?

Posted by jaleo - 2008/03/09 23:38

Thanks,Steve

It works now.

This is a great component.

Re:utf8 support?

Posted by Steve - 2008/03/10 18:31

Thanks,Steve

It works now.

This is a great component.

Glad to hear you got it working jaleo . :)

montemedia

I have made changes to the text file.

Try removing everything in `function numeric_html($text){}` so it looks like this:

```
function numeric_html($text){
return $text;
}
```

And then just replace this with the code in the text file:

```
/******
```

```
High end numeric converter
```

```
*****/
```

```
function high_order($text){
```

```
$text = preg_replace('/(.)se', "&# . ((ord(substr('$1', 0, 1)) - 192) * 64 + (ord(substr('$1', 1, 1)) - 128)) . ';"", $text);
```

```
$text = preg_replace('/(..)se', "&# . ((ord(substr('$1', 0, 1)) - 224) * 4096 + (ord(substr('$1', 1, 1)) - 128) * 64 + (ord(substr('$1', 2, 1)) - 128)) . ';"", $text);
```

```
return $text;
```

```
}
```

If this is still no good it shows where to remove code in the text file.

```
=====
```