
Issues with Special Characters In Imported Content

Posted by carly23420 - 2008/02/24 19:44

Hi,

Has anyone else had the problem that once the content has been imported, it shows special characters - such as quotation marks and apostrophes - as a symbol, such as this:

I'm loving this component, but am really not looking forward to having to manually correct more than 1,000 content items by replacing symbols with quotation marks and apostrophes, etc.

If there's a fix or workaround, I'd appreciate any thoughts!

Thanks -

Carly

Re:Issues with Special Characters In Imported Content

Posted by Steve - 2008/02/24 20:09

Hi Carly

What version of Export Content are you using?
What versions of Joomla are you attempting to move content to?

Have you copy and pasted Ms word docs into Joomla?

Re:Issues with Special Characters In Imported Content

Posted by carly23420 - 2008/02/24 21:00

I'm using Joomla 1.0.15 and Export Content 1.7.0... I haven't copied and pasted any Word documents; I always use notepad or something similar without formatting.

I just thought of something, though... The new site (where I'm importing) is running Joomla 1.0.15, and the old site (where the exported content is coming from) is running Joomla 1.0.13. Could that be causing the problem?

Re:Issues with Special Characters In Imported Content

Posted by Steve - 2008/02/24 21:17

It could be this is an encoding issue.

Export content is set to move content over to utf8 from iso-8859-1 and should make all the changes needed.

Anyway here is a way to fix things up.
First of all I would backup the entire data base.
Then if you fire up phpMyAdmin and enter this you should be able to tidy things up in a couple of minutes.

```
update `jos_content` set `introtex` = replace(`introtex`, 'replace_that', 'with_this');  
update `jos_content` set `fulltext` = replace(`fulltext`, 'replace_that', 'with_this');
```

Where "replace_that" is the strange character "with_this" is the replacement.
Just run this for each strange character you find.

I would suggest trying to use standard punctuation instead of ms word punctuations where ever possible.

If your not sure how to go about it please post back here and I will walk you through it.

Re:Issues with Special Characters In Imported Content

Posted by guarez - 2008/02/27 18:09

Hi.. i have the same error with my exported database from my old mambo 4.5.2 file.

When i install the module (files with the data) in my new joomla 1.5 i get this error

JInstaller::install: SQL Error. DB function failed with error number 1064

You have an error in your SQL syntax;

check the manual that corresponds to your MySQL server version for the right syntax to use near 'segundo a~3e cancelo la mayor fiesta electronica del mundo.

Aqui las

razones y' at line 1 SQL=INSERT INTO jos_export_content (`id`, `title`, `alias`, `title_alias`,

`introtext`, `fulltext`, `state`, `sectionid`, `mask`,

`catid`, `created`, `created_by`, `created_by_alias`, `modified`, `modified_by`, `checked_out`, `checked_out_time`,

`publish_up`, `publish_down`,

`images`, `urls`,

`attribs`, `version`, `parentid`, `ordering`, `metakey`, `metadesc`,

`access`, `hits`, `metadata`)VALUES('68','Mientras Chile vive un boom,

Love Parade no va en Berl','','Por segundo a~3e cancelo la mayor fiesta electronica del mundo.

Aqui las razones y el futuro del desfile en Santiago.')

Loveparade 2006

PAULINA TORO

El desfile del amor esta fuera de control. Unas dos millones de personas se esperaban para Love Parade este a~%n las calles de Berlin.

Pero otra vez nada.

Igual que en 2004, no habra Love Parade en Alemania.

No hubo ni permisos ni el auspicio necesarios para hacerla.

No hubo tiempo tampoco para reunir ambas variables. No hubo organizacion.

Love Parade esta colapsada.

Desde las oficinas de Love Parade Berlin, el comunicado

que avisa su cancelacion es confuso: "

SQL =

INSERT INTO jos_export_content (`id`, `title`, `alias`, `title_alias`,

`introtext`, `fulltext`, `state`, `sectionid`, `mask`, `catid`, `created`,

`created_by`, `created_by_alias`, `modified`, `modified_by`, `checked_out`, `checked_out_time`, `publish_up`,

`publish_down`, `images`, `urls`,

`attribs`, `version`, `parentid`, `ordering`, `metakey`, `metadesc`,

`access`, `hits`, `metadata`)VALUES('68','Mientras Chile vive un boom,

Love Parade no va en Berl','','Por segundo a~3e cancelo la mayor fiesta

electronica del mundo. Aqui las razones y el futuro del desfile en Santiago.')

'Loveparade 2006PAULINA TOROEI desfile del amor esta fuera de control.

Unas dos millones de personas se esperaban para Love

Parade este a~%n las calles de Berlin.

Pero otra vez nada. Igual que en 2004, no habra Love Parade en Alemania.

No hubo ni permisos ni el auspicio necesarios para hacerla.

No hubo tiempo tampoco para reunir ambas variables. No hubo organizacion.

Love Parade esta colapsada.

Desde las oficinas de Love Parade Berlin, el comunicado que avisa

su cancelacion es confuso:

"Component Install: SQL error or missing or unreadable SQL file.

DB function failed with error number 1064

You have an error in your SQL syntax; check the manual that

corresponds to your MySQL server version for the right syntax to use

near 'segundo a~3e cancelo la mayor fiesta electronica del mundo.
Aqui las razones y' at line 1 SQL=INSERT INTO jos_export_content (`id`, `title`, `alias`, `title_alias`,
`introtex`, `fulltext`, `state`, `sectionid`,
`mask`, `catid`, `created`, `created_by`, `created_by_alias`,
`modified`,
`modified_by`, `checked_out`, `checked_out_time`, `publish_up`,
`publish_down`, `images`, `urls`, `attribs`, `version`, `parentid`,
`ordering`, `metakey`, `metadesc`, `access`, `hits`, `metadata`)
VALUES('68','Mientras Chile vive un boom, Love Parade no va en Berl',
",", 'Por segundo a~3e cancelo la mayor fiesta electronica del mundo.
Aqui las razones y el futuro del desfile en Santiago.',
Loveparade 2006
PAULINA TORO

El desfile del amor esta fuera de control. Unas dos millones de personas se
esperaban para Love Parade este a~%n las calles de Berlin.
Pero otra vez nada. Igual que en 2004,
no habra Love Parade en Alemania.
No hubo ni permisos ni el auspicio necesarios para hacerla.
No hubo tiempo tampoco para reunir ambas variables.
No hubo organizacion. Love Parade esta colapsada.

Desde las oficinas de Love Parade Berlin, el comunicado que avisa su cancelacion es confuso: "
SQL =
INSERT INTO jos_export_content (`id`, `title`, `alias`, `title_alias`, `introtex`,
`fulltext`, `state`, `sectionid`, `mask`, `catid`, `created`, `created_by`, `created_by_alias`,
`modified`, `modified_by`, `checked_out`,
`checked_out_time`, `publish_up`, `publish_down`, `images`, `urls`, `attribs`, `version`,
`parentid`, `ordering`, `metakey`, `metadesc`, `access`, `hits`, `metadata`)
VALUES('68','Mientras Chile vive un boom, Love Parade no va en Berl',",",
Por segundo a~3e cancelo la mayor fiesta electronica del mundo.
Aqui las razones y el futuro del desfile en Santiago.',
Loveparade 2006PAULINA TOROEI desfile del amor esta fuera de control.
Unas dos millones de personas se esperaban para
Love Parade este a~%n las calles de Berlin.
Pero otra vez nada.
Igual que en 2004,
no habra Love Parade en Alemania.
No hubo ni permisos ni el auspicio necesarios para hacerla.
No hubo tiempo tampoco para reunir ambas variables.
No hubo organizacion.
Love Parade esta colapsada.Desde las oficinas de Love Parade Berlin,
el comunicado que avisa su cancelacion es confuso: "

i convert all my special chars (ñ,á,é,í etc..) to normal characters..
but i can impor my data =/

any help?

=====

Re:Issues with Special Characters In Imported Content

Posted by Steve - 2008/02/28 18:36

So to be clear you have Export Content version 1.0.0 installed on Mambo version 4.5.2
And Export Content version 2.0.0 installed on Joomla 1.5

You then compile the data on the Mambo site using the "compile for Joomla 1.5" button

Then when attempting to install on the Joomla 1.5 site as a component you get the error.

If so I think you will find the "~" you have in the text is causing the error and also I see "%" as well.

These symbols are used by mysql as operators and should have been converted when compiling the data.

I will look into fixing this in the next release for now though you can use the above method on the Mambo site before compiling.

Or on the Mambo site you could go to /administrator/components/com_export_content/convert_img.php. Scroll down to about line 2186 where you will see:

```
$scr=array(  
Change it to:
```

```
$scr=array('~' => '& # 126;', '%'=> '& # 37;',  
save and re-compile.
```

Note there should be no spaces between the "& #" or 126 or 37 in the above code but it is the only way to get it to display here on the forums.

Here is a table showing how they should look under the "Numerical Code" column as the forum displays them incorrectly. http://www.bestdownloadsites.com/ascii_table.html

Re:Issues with Special Characters In Imported Content

Posted by carly23420 - 2008/03/07 17:27

Hi Steve,

An update: I was unable to use your suggestion for the SQL commands on the new site, as the odd character was showing up for multiple instances (dashes, question marks, ellipses, an array of things). What I did do, however, was use your suggestion to track down improperly encoded characters from the original site (the site where I exported the content) and then change them to the correctly encoded format. I then re-exported all of the data, and reimported everything again on the new site. Voila! It worked.

Thanks so much for your suggestion. While it did take me a while to get through it all, it was much less of a headache than it would have been to copyedit each imported document by hand.

I appreciate your help!

Carly

Re:Issues with Special Characters In Imported Content

Posted by tlearyus - 2008/03/10 01:12

i have the same problem with importing from/into joomla 1.0.13 and 1.0.14

example

There is no need to establish a community recreation node around Torakina Park it is already happening. Plans in this precinct for commercial development near the surf club and accommodation on land south of the club are also worrying.

Re:Issues with Special Characters In Imported Content

Posted by Steve - 2008/03/11 18:08

Hi tlearyus

Export Content seems to have picked some control character or ms word.

I have attached a patch that should sort out the problem.

Download it locally and unzip it then replace:
/administrator/components/com_export_content/convert_img.php with the new one.
Then re-compile and see how it goes. :)
http://www.bestdownloadsites.com/export_content/images/fbfiles/files/convert_img.zip

=====

Re:Issues with Special Characters In Imported Cont

Posted by tlearyus - 2008/03/13 21:33

many thanks - really appreciate it !!

will let you know how it goes.

cheers - tlearyus

=====

Re:Issues with Special Characters In Imported Cont

Posted by tlearyus - 2008/03/13 23:39

sorry but it didn't work, added the updated file to my current Joomla setup and did an export

then did a fresh joomla 1.0.14 install on a new host, installed the latest SVN release of Joomla export and imported the content by sections and still getting all the strange chars.

would love to get this working so let me know what you need to help fix this.

cheers - glen

=====

Re:Issues with Special Characters In Imported Cont

Posted by tlearyus - 2008/03/13 23:59

ooops - double post.

=====

Re:Issues with Special Characters In Imported Content

Posted by Steve - 2008/03/14 17:43

Hi Glen

i have the same problem with importing from/into joomla 1.0.13 and 1.0.14

Please give me more details to make sure we are on the same page.

Are you trying to move content between to sites from the Joomla 1.0+ series like 1.0.13 for example or from 1.0+ to 1.5+.

If so what language are the sites in.

If they are Latin based (English for example) and you are moving between the 1.0+ Joomla series you may be able to use the old Export Content version 1.6

You may also find this thread useful.

http://www.bestdownloadsites.com/export_content/component/option,com_fireboard/Itemid,13/func,view/id,926/catid,1/

=====

Re:Issues with Special Characters In Imported Cont

Posted by copiermad - 2008/03/27 08:54

Hi,

Great comments on this component. Although my encoding problems do not appear to be going away.

My setup: -

From:
Joomla 1.0.14
Export Content 1.7

To:
Joomla 1.0.14
Export Content 1.7

I am exporting a section that has items containing amongst other things like apostrophes (').

Without any changes to the component:
The content component installs ok, but the code gets re-written to control chars as well as other stuff going wrong.

When your patch is applied:
The content component will not install, due to control characters in the content.

When removing all code from the numeric_html & high_order functions to return \$text unchanged:
The content component will not install, due to control characters in the content.

Is there a way to get this component to transfer the content "as is", so I do not have to edit the content the other end? I am happy to make changes to the code and will even attempt to re-write the component if necessary.

Just as a matter of course, would the whole issue of encoding go away if you just base64 encoded the whole lot before putting it in the xml file. Then base64 decode the whole lot on extraction?

Thanks,
Copiermad :)

=====

Re:Issues with Special Characters In Imported Cont

Posted by Steve - 2008/03/27 21:49

Hi Copiermad.

Is there a way to get this component to transfer the content "as is", so I do not have to edit the content the other end? I am happy to make changes to the code and will even attempt to re-write the component if necessary.

Just as a matter of course, would the whole issue of encoding go away if you just `_base_64` encoded the whole lot before putting it in the `_xml_` file. Then `_base_64` decode the whole lot on extraction?

I have been thinking about by passing the XML by writing all the data to a separate file that may be executed after the install process this would do away with the whole issue of trying to parse characters via XML.

The only thing is finding the time as I'm very busy with another large paying project.
If you want to give it a go by all feel free and please post back here so it can be shared with the rest of the community.

Your idea about `base_64` encoding may be the way to go for you at the moment to get out of trouble the only issue is that it will make the upload size larger and therefore restrict how much can be moved at one time other than that I see no problem with it.

If you want to give it a go the best way to go about it would be to change:

administrator/components/com_export_content/convert_img.php

function numeric_html(\$text){} to do the encoding.

And in /administrator/components/com_export_content/admin.export_content.

About line 2000 you will see:

```
function multiSectionSave( $option, $cid) {}
```

This is where after the content has been installed and you want to transfer by sections.

As you scroll down you will see that there is a decoding function this is where you could do the base_64 decoding:

```
$sect_title = decode_entities($sect->title);
```

The decode_entities() function is about line 2500 of /administrator/components/com_export_content/convert_img.php

One other thing I would remove everything in the high_order function as instructed earlier in this thread.

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

```
    return $text;
```

```
}
```

Please post back with any questions. :)

Re:Issues with Special Characters In Imported Content

Posted by Steve - 2008/04/02 18:53

Another thread that may help.

Changing encoding
