

Problems with UTF-8 encoding



Felipe Cypriano 37 posts since

Mar 18, 2009

Hello,

I'm having a annoying problem. I use IntelliJ IDEA 8.1.1 #9805 and my project is configured to use UTF-8 encoding. The problem is, I made a page (using zk framework, bu I don't it matters) to build a tree for a menu:

And the first item of the tree is:

```
<treerow>
  <treecell label="Página Inicial"/>
</treerow>
```

This was working and the "á" character appears correct on the browser. But I need to to build the menu from the code to it be dynamic, so I did a java file with this:

```
static {
    menuItems = new ArrayList();
    menuItems.add("Página Inicial");
}
```

And bound the page to this menuItems, and now when the page is loaded in the browser I get this: "PÃ¡gina Inicial". The meta content type is set correctly to UTF-8.

It only works if I change it to menuItems.add("P\u00E1gina Inicial"), but shouldn't it be automacally done by the IDE?

Regards,

Felipe Marin Cypriano

Tags: utf-8



[Alexander Chernikov](#) 567 posts since

Jan 11, 2007 1. **Re: Problems with UTF-8 encoding** Apr 16, 2009 6:23 PM

Hello.

1. When your java file is opened, is "UTF-8" shown in the status bar?
2. Could you check what is the real encoding of the java source file (in some external editor, or anywhere).
3. I'm not familiar with zk. How java source is transformed into a browser page: is the source compiled, or html is generated directly from source, or ... ?

Alexander.



[Felipe Cypriano](#) 37 posts since

Mar 18, 2009 2. **Re: Problems with UTF-8 encoding** Apr 16, 2009 7:50 PM

1. When your java file is opened, is "UTF-8" shown in the status bar?

Yes, the status bar shows "UTF-8" and it's disabled I can't change (I think it's because the code has string literals).

2. Could you check what is the real encoding of the java source file (in some external editor, or anywhere).

I've opened the file using Notepad++ in the status it shows: Dos\Windows | ANSI as UTF-8

3. I'm not familiar with zk. How java source is transformed into a browser page: is the source compiled, or html is generated directly from source, or ... ?

I'm pretty new to ZK, but the java source is kind a controller. In that case I've a zul page (xml file) which get the menu items from the compiled java source and render the html. Is that what you want to know?



[Alexander Chernikov](#) 567 posts since

Jan 11, 2007 3. **Re: Problems with UTF-8 encoding** Apr 17, 2009 7:14 PM

Thank you. My 3rd question was to understand, who translates "proper" java into "unproper" html. It seems we should suspect IDEA compiler. I tried simple application in 9814 (I believe, in 9805 the behavior was the same):

```
public class Speaker {  
    public static void main(final String[] argv) {  
        System.out.println("National chars here.");  
    }  
}
```

The source file is in UTF-8.

In 2 cases of IDEA settings combinations:

1. IDE encoding = UTF-8, project encoding = (not set).
2. IDE encoding = system default, project encoding = UTF-8.

Compiler produces proper .class file: national chars are printed correctly.

Is it the same for you?

Alexander.



[Felipe Cypriano](#) 37 posts since

Mar 18, 2009 4. **Re: Problems with UTF-8 encoding** Apr 17, 2009 7:39 PM

Problems with UTF-8 encoding

Cases that worked for me, using build #9805:

1. IDE encoding = UTF-8, project encoding = not set (but the file is UTF-8)
2. IDE encoding = UTF-8, project encoding = UTF-8

Those doesn't work:

1. IDE encoding = UTF-8, project encoding = not set (but the file is windows-1252, which is the system default)
2. IDE encoding = not set (<< Systems Default>>), project encoding = UTF-8



[Alexander Chernikov](#) 567 posts since

Jan 11, 2007 5. **Re: Problems with UTF-8 encoding** Apr 17, 2009 8:32 PM

> Those doesn't work:

> 1. IDE encoding = UTF-8, project encoding = not set (but the file is windows-1252, which is the system default)

This is correct, since a source in windows-1252 is compiled as UTF-8.

> 2. IDE encoding = not set (<< Systems Default>>), project encoding = UTF-8

And the file encoding is ... ?

Actually the rule is simple: make project encoding equal to file encoding, and compiler should follow. IDE encoding is used when project one is not set.

Alexander.



[Felipe Cypriano](#) 37 posts since

Problems with UTF-8 encoding

Mar 18, 2009 6. **Re: Problems with UTF-8 encoding** Apr 17, 2009 8:59 PM

> 2. IDE encoding = not set (<< Systems Default>>), project encoding = UTF-8

And the file encoding is ... ?

The file encoding is UTF-8, just like the project. And with this config the result isn't what I expected.



[Alexander Chernikov](#) 567 posts since

Jan 11, 2007 7. **Re: Problems with UTF-8 encoding** Apr 17, 2009 9:20 PM

This is certainly a bug. The problem is that I cannot reproduce it.



[Felipe Cypriano](#) 37 posts since

Mar 18, 2009 8. **Re: Problems with UTF-8 encoding** Apr 19, 2009 1:11 AM

Now I tested it using build #9815 hoping the problem was solved but this build behaves the same way as #9805.

I've created a simple project from scratch to test, see the attached screenshot.

My environment is

- IntelliJ IDEA (Diana) #9815
- JDK 1.6.0_07
- Windows Vista Business 32 bit with SP1 (if it matters: Brazilian version, portuguese as default language and windows-1252 or ISO-8859-1 as system default encoding)

Attachments:



[Bug UTF-8.JPG](#) (226.7 K)



[Felipe Cypriano](#) 37 posts since

Mar 18, 2009 9. **Re: Problems with UTF-8 encoding** Apr 17, 2009 8:02 PM

Problems with UTF-8 encoding

One more thing, if I execute the generated class from the command line the characters aren't correctly displayed (even when they are in IDEA console):

```
public class TestUTF8 {  
    public static void main(String[] args) {  
        System.out.println("ç á ò ã");  
    }  
}
```

Execute in the command line (cmd):

```
C:> java TestUTF8  
p ß _ Ö
```



[Alexander Chernikov](#) 567 posts since

Jan 11, 2007 10. **Re: Problems with UTF-8 encoding** Apr 17, 2009 8:28 PM

The issue with console is possible due to improper cmd code page and not true type console font.

You can try to compile UTF-8 source from command line and run it there: I think the effect will also be not desirable.

Alexander.



[Viraf Karai](#) 71 posts since

Feb 28, 2009 11. **Re: Problems with UTF-8 encoding** Apr 19, 2009 4:19 AM

So it looks like you're okay inside IntelliJ, but just to clarify, you can set the encoding for your IntelliJ project by selecting ctrl-alt-s and then selecting 'File Encodings' in the left panel (see attached screenshot).

As far as the command line, goes, try running your code as follows:

Problems with UTF-8 encoding

~~~~~

```
$ java -Dfile.encoding=UTF-8 <yourClassname>
```

and let us know if this makes any difference

**Attachments:**

- 



[Screenshot-Settings.png](#) (34.0 K)