

## µWeb - Bug #626

### utf8 in conditions in templateparser goes awry

2012-02-20 11:36 - Jan Klopper

<b>Status:</b>	Closed	<b>Start date:</b>	2012-02-20
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Elmer de Loeff	<b>% Done:</b>	100%
<b>Category:</b>	TemplateParser	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>		<b>Spent time:</b>	1.75 hour
<b>Description</b>			
<p>&lt;a href="" class="button"&gt;↑&lt;/a&gt; is handled just fine, however:</p> <pre>{{ if [paragraph:sort] &gt; 0 }} &lt;a href="" class="button"&gt;↑&lt;/a&gt; {{ endif }}</pre>			
<p>Gives me an error:</p> <pre>Traceback (most recent call last):   File "/usr/lib/python2.6/dist-packages/mod_python/importer.py", line 1537, in HandlerDispatch     default=default_handler, arg=req, silent=hlist.silent)   File "/usr/lib/python2.6/dist-packages/mod_python/importer.py", line 1229, in _process_target     result = _execute_target(config, req, object, arg)   File "/usr/lib/python2.6/dist-packages/mod_python/importer.py", line 1128, in _execute_target     result = object(arg)   File "/home/underdark/underdark/libs/uweb/__init__.py", line 110, in RequestHandler     response = pages.InternalServerError(*sys.exc_info())   File "/home/underdark/underdark/libs/uweb/pagemaker/__init__.py", line 329, in InternalServerErr or     'traceback': self._ParseStackFrames(traceback)))   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 263, in Parse     return SafeString(''.join(tag.Parse(**kws) for tag in self))   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 263, in &lt;genexpr&gt;     return SafeString(''.join(tag.Parse(**kws) for tag in self))   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 459, in Parse     output.append(''.join(tag.Parse(**replacements) for tag in self))   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 459, in &lt;genexpr&gt;     output.append(''.join(tag.Parse(**replacements) for tag in self))   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 459, in Parse     output.append(''.join(tag.Parse(**replacements) for tag in self))   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 459, in &lt;genexpr&gt;     output.append(''.join(tag.Parse(**replacements) for tag in self))   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 564, in Parse     value = TAG_FUNCTIONS[func](value)   File "/home/underdark/underdark/libs/uweb/templateparser.py", line 632, in HtmlEscape     text = unicode(text) UnicodeDecodeError: 'ascii' codec can't decode byte 0xe2 in position 500: ordinal not in range(128 )</pre>			

#### Associated revisions

##### Revision 3439:403be453c6ae - 2012-02-20 12:53 - Elmer de Loeff

Added tests to the TemplateParser suite to test for Unicode support in templates. They may now contain raw codepoints, and these are converted to UTF8. This resolves #626.

##### Revision 188:55f11f246503 - 2012-02-20 12:53 - Elmer de Loeff

Added tests to the TemplateParser suite to test for Unicode support in templates. They may now contain raw codepoints, and these are converted to

UTF8. This resolves #626.

## History

---

### #1 - 2012-02-20 11:45 - Jan Kloppe

Hmm, untrue,

It appears it works fine if I use the correct variable.

However.

I would expect an index error / instead of an encode error if 'sort' doesn't exist.

### #2 - 2012-02-20 11:49 - Elmer de Looff

- Description updated

### #3 - 2012-02-20 12:11 - Elmer de Looff

- Status changed from New to In Progress

- % Done changed from 0 to 30

That's a very weird error. These are my results when doing some quick checks (there definitely is an error here):

Encoded UTF8 in the template text (that is, not in a replacement tag, as in your example):

```
>>> import templateparser
>>> tpl = u'We \u2665 Python'
>>> templateparser.Template(tpl.encode('utf8')).Parse()
'We \xe2\x99\xa5 Python'
```

This works, as expected. The raw bytestream that goes in, is sent raw to the output. Unicode codepoint U2665 translates correctly to the 3-byte UTF8 string. Parsing the same template without first encoding the template to UTF8 (or any other bytestream, though that will result in character set mismatches, and is a **very** bad idea); that is, storing raw Unicode in the template, does result in an error:

```
>>> import templateparser
>>> tpl = u'We \u2665 Python'
>>> templateparser.Template(tpl).Parse()
Traceback (most recent call last):
  File "<stdin>", line 1, in <module>
  File "/home/elmer/underdark/libs/uweb/templateparser.py", line 248, in __init__
    self.AddString(raw_template)
  File "/home/elmer/underdark/libs/uweb/templateparser.py", line 308, in AddString
    self._ExtendText(node)
  File "/home/elmer/underdark/libs/uweb/templateparser.py", line 365, in _ExtendText
    for node in self.TagSplit(node):
  File "/home/elmer/underdark/libs/uweb/templateparser.py", line 329, in TagSplit
    yield TemplateText(node)
UnicodeEncodeError: 'ascii' codec can't encode character u'\u2665' in position 3: ordinal not in range(128)
```

The error here is what we expect, the Unicode codepoint is *not* in the ASCII character set. This is a shortcoming of TemplateParser and I'll look into fixing this.

Back to your problem, the reported error mentions `\xe2`. This is likely not to be a Unicode codepoint (the associated character is `Î`), but the first character of the up-arrow in UTF8: `\xe2\x86\x91`.

What I suspect goes wrong here is that when requesting the `:sort` index, you get the `sort` method of whatever object you're working on. If the representation of this object contains raw Unicode, the `templateparser` barfs on this. However, this does still not explain the exact error given, because that does not contain a Unicode codepoint, but the first byte of a multi-byte sequence.

```
>>> import templateparser
```

```
>>> tpl = '[obj:sort]'  
>>> tp.Template(tpl.encode('utf8')).Parse(obj=[])  
'&lt;built-in method sort of list object at 0x273f998&gt;'
```

#### #4 - 2012-02-20 15:20 - Elmer de Looff

- Status changed from *In Progress* to *Resolved*
- % Done changed from 30 to 70

Applied in changeset commit:403be453c6ae.

#### #5 - 2012-02-20 15:52 - Elmer de Looff

- Status changed from *Resolved* to *Feedback*
- Assignee changed from *Elmer de Looff* to *Jan Klopper*
- Priority changed from *High* to *Normal*

Jan,

The main bug has been solved. Please see if you can reproduce the bug you encountered earlier, I'd like to know what causes it.

#### #6 - 2012-02-21 11:33 - Elmer de Looff

- Status changed from *Feedback* to *Resolved*

Applied in changeset commit:138c5a8a1c0e.

#### #7 - 2012-02-22 17:19 - Elmer de Looff

Applied in changeset [55f11f246503](#).

#### #8 - 2012-02-22 17:19 - Elmer de Looff

Applied in changeset [55f11f246503](#).

#### #9 - 2012-02-28 11:19 - Jan Klopper

- Assignee changed from *Jan Klopper* to *Elmer de Looff*

When triggering this bug now, i get the following error:

```
<class 'underdark.libs.uweb.templateparser.TemplateKeyError'>
```

Item has no index, key or attribute 'sort'.

Seems to work as expected, thnx!

#### #10 - 2012-02-28 11:21 - Elmer de Looff

- Status changed from *Resolved* to *Closed*
- % Done changed from 70 to 100

In that case, no problem remains (there might be one hidden somewhere, but without any way to test this, I'll close it for now).

**#11 - 2012-02-28 11:21 - Elmer de Looff**

- *Category set to TemplateParser*