

## Redmine - Defect #20730

### Fix tokenization of phrases with non-ascii chars

2015-09-13 05:56 - Jens Krämer

<b>Status:</b>	Closed	<b>Start date:</b>	
<b>Priority:</b>	Normal	<b>Due date:</b>	
<b>Assignee:</b>	Jean-Philippe Lang	<b>% Done:</b>	0%
<b>Category:</b>	Search engine	<b>Estimated time:</b>	0.00 hour
<b>Target version:</b>	3.0.6	<b>Affected version:</b>	
<b>Resolution:</b>	Fixed		
<b>Description</b>			
\w only matches ASCII characters, we should either use [:alnum:] instead or simply match all non-" characters for the phrase. Test case included.			

#### Associated revisions

##### Revision 14662 - 2015-10-09 09:35 - Jean-Philippe Lang

Fix tokenization of phrases with non-ascii chars (#20730).

Patch by Jens Krämer.

##### Revision 14663 - 2015-10-09 09:36 - Jean-Philippe Lang

Merged r14662 (#20730).

##### Revision 14664 - 2015-10-09 09:36 - Jean-Philippe Lang

Merged r14662 (#20730).

#### History

##### #1 - 2015-09-27 17:30 - Go MAEDA

- Tracker changed from Patch to Defect

- Target version set to 3.1.2

+1

Search keyword "テスト" (written in Japanese) matches both "テスト" and "テスト" in the current trunk, but it should not match the latter.

##### expected:

```
Redmine::Search::Fetcher.new("テスト", ...).tokens => ["テスト"]
```

##### actual:

Redmine::Search::Fetcher.new("テスト", ...).tokens => ['テスト']

This behavior can be fixed by this patch.

**#2 - 2015-10-09 09:37 - Jean-Philippe Lang**

- Status changed from New to Closed
- Assignee set to Jean-Philippe Lang
- Target version changed from 3.1.2 to 3.0.6
- Resolution set to Fixed

Patch applied, thanks.

**Files**

---

fix-tokenization-for-phrases-with-non-ascii-characte.patch	1.39 KB	2015-09-13	Jens Krämer
--	---------	------------	-------------