

Redmine - Defect #20730

Fix tokenization of phrases with non-ascii chars

2015-09-13 05:56 - Jens Krämer

Status:	Closed	Start date:	
Priority:	Normal	Due date:	
Assignee:	Jean-Philippe Lang	% Done:	0%
Category:	Search engine	Estimated time:	0.00 hour
Target version:	3.0.6	Affected version:	
Resolution:	Fixed		
Description			
\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 "検索 検索" (written in Japanese) matches both "検索 検索" and "検索検索" in the current trunk, but it should not match the latter.

expected:

```
Redmine::Search::Fetcher.new('検索 検索 ', ...).tokens => ['検索 検索 ']
```

actual:

```
Redmine::Search::Fetcher.new('検索 検索 ', ...).tokens => ['検索 ', '検索 ']
```

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
--	---------	------------	-------------