Fix tokenization of phrases with non-ascii chars

<table>
<thead>
<tr>
<th>Status:</th>
<th>Closed</th>
<th>Start date:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Priority:</td>
<td>Normal</td>
<td>Due date:</td>
</tr>
<tr>
<td>Assignee:</td>
<td>Jean-Philippe Lang</td>
<td>% Done: 0%</td>
</tr>
<tr>
<td>Category:</td>
<td>Search engine</td>
<td>Estimated time: 0.00 hour</td>
</tr>
<tr>
<td>Target version:</td>
<td>3.0.6</td>
<td></td>
</tr>
<tr>
<td>Resolution:</td>
<td>Fixed</td>
<td></td>
</tr>
</tbody>
</table>

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

Search keyword “検証 テスト*” (written in Japanese) matches both 検証 テスト* and 検証テスト* in the current trunk, but it should not match the latter.

expected:

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

actual:
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

<table>
<thead>
<tr>
<th>Name</th>
<th>Size</th>
<th>Date</th>
<th>Author</th>
</tr>
</thead>
<tbody>
<tr>
<td>fix-tokenization-for-phrases-with-non-asci-characte.patch</td>
<td>1.39 KB</td>
<td>2015-09-13</td>
<td>Jens Krämer</td>
</tr>
</tbody>
</table>