

Redmine - Patch #30037

Allow single Chinese character as a search keyword

2018-11-27 06:04 - Go MAEDA

Status:	Closed	Start date:	
Priority:	Normal	Due date:	
Assignee:	Go MAEDA	% Done:	0%
Category:	Search engine	Estimated time:	0.00 hour
Target version:	4.0.0		
Description			
<p>Currently, Redmine requires at least 2 characters length for search keywords. It is not a problem for languages such as English and French.</p> <p>But for languages such as Japanese and Chinese, the limitation imposes inconvenience for users because there are some words which consist of only a single character. Some examples as follows:</p> <ul style="list-style-type: none">• 金 = money in Japanese and Chinese• 水 = water in Japanese and Chinese• 野猪 = a wild boar in Japanese, a pig in Chinese• 王 = a common family name for Chinese people <p>To allow such single character search keywords, I suggest the following patch. It accepts single multibyte-character keywords while keeping the current limitation for keywords with ASCII characters.</p> <pre>diff --git a/lib/redmine/search.rb b/lib/redmine/search.rb index 674022151..d4c0b9b20 100644 --- a/lib/redmine/search.rb +++ b/lib/redmine/search.rb @@ -59,8 +59,8 @@ module Redmine # extract tokens from the question # eg. hello "bye bye" => ["hello", "bye bye"] @tokens = @question.scan(%r{((\s ^)"[^"]+(\s \$) \S+)}).collect { m m.first.gsub(%r{(^\s *" \s* \s* " \s*\$)}, '')} - # tokens must be at least 2 characters long - @tokens = @tokens.uniq.select { w w.length > 1 } + # tokens must be at least 2 characters long in ASCII + @tokens = @tokens.uniq.select { w w.bytesize > 1 } # no more than 5 tokens to search for @tokens.slice! 5..-1 end</pre>			

Associated revisions

Revision 17667 - 2018-11-30 16:42 - Go MAEDA

Allow single Chinese character as a search keyword (#30037).

Patch by Go MAEDA.

History

#1 - 2018-11-27 11:00 - Go MAEDA

- File 30037-allow-single-multibyte-search-keyword.diff added

Added a test.

#2 - 2018-11-27 12:34 - Tomohisa Kusakawa

+1

#3 - 2018-11-27 17:26 - Anonymous

+0⁰

#4 - 2018-11-28 06:16 - Go MAEDA

- Target version set to 4.1.0

I am sure that this is a necessary improvement for people who speak Japanese and Chinese because there are some words expressed by only a single character as I wrote in the description field. They cannot search such word with the current versions of Redmine.

Setting the target version to 4.1.0.

#5 - 2018-11-30 04:53 - Go MAEDA

- File 30037-allow-single-chinese-character-search-keywords.diff added

- Subject changed from Allow single multibyte character as a keyword in search to Allow single character search keywords for Chinese characters

Updated the patch. The new patch accepts single character tokens only if they are a Chinese character (□□/□□).

The previous patch accepts all multibyte characters. But the implementation has some problems. First, accepting only Chinese characters is enough. Even in CJK languages, it is not necessary to search single characters other than Chinese characters such as Japanese Kana (□□) and Korean Hangul (□□) because they are syllabic writing systems and single character of those are not meaningful words in most cases. Second, the previous patch also accepts single latin characters such as "á", "ó", and "ß". I think it is a regression. Current versions of Redmine does not accept those single characters.

In order to solve these issues, the new patch allows character tokens only for Chinese characters and keeps the current behavior for other characters.

#6 - 2018-11-30 16:42 - Go MAEDA

- Subject changed from Allow single character search keywords for Chinese characters to Allow single Chinese character as a search keyword

- Status changed from New to Closed

- Assignee set to Go MAEDA

- Target version changed from 4.1.0 to 4.0.0

Committed.

Files

30037-allow-single-multibyte-search-keyword.diff	1.35 KB	2018-11-27	Go MAEDA
30037-allow-single-chinese-character-search-keywords.diff	1.26 KB	2018-11-30	Go MAEDA