redmineorg-copy202205 - Vote #79660

Allow single Chinese character as a search keyword

2022/05/09 18:39 - Admin Redmine

	:	Closed		:	2022/05/09
	:			:	
	:			:	0%
	:	Search engine_16		:	0.00
	:	4.0.0_99		:	0.00
Redmineorg_URL:		https://www.redmine.org/issues/30037	**status_id:**		5
category_id:		16	**tracker_id:**		3
version_id:		99	**plus1:**		1
issue_org_id:		30037	**affected_version:**		
author_id:		332	**closed_on:**		
assigned_to_id:		332	**affected_version_id:**		
comments:		6			

Currently, Redmine requires at least 2 characters length for search keywords. It is not a problem for languages such as English and French.

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:

- = money in Japanese and Chinese
- = water in Japanese and Chinese
- = a wild boar in Japanese, a pig in Chinese
- = a common family name for Chinese people

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.

```
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
```

journals

Added a test.

+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.

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", "　", 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.

Committed.

#1 - 2022/05/10 17:04 - Admin Redmine

- 　　　　　　*Search engine_16*

- 　　　　　　　*4.0.0_99*