

Redmine - Defect #6551

Highlighting in search results is case sensitive for cyrillic pattern

2010-10-01 09:57 - Alexey Ivlev

Status:	New	Start date:	2010-10-01
Priority:	Normal	Due date:	
Assignee:		% Done:	50%
Category:	Search engine	Estimated time:	0.00 hour
Target version:		Affected version:	1.0.1
Resolution:			
Description			
<p>I am sorry for my persistence, I have published the same problem on forum, but still have no response from there...</p> <p>When I search any pattern in english everything works fine - highlighting in search results is case insensitive. If I try to search pattern in russian I have case insensitive search output, but highlighting my pattern in that results is case sensitive.</p> <p>For example, if I try to search "ам" (all in lowercase) pattern I will see the next:</p> <p>ам - all letter in search output are lowercase and highlighting work fine</p> <p>Ами, - but if it is one letter or more is uppercase, highlighting doesn't appear.</p> <p>In source code of search results page this tag <code>ам</code> exist for first image and does not exists for the second.</p> <p>I use MySQL 5.1.41 database with utf8_general_ci encoding and apache + passenger on Ubuntu 10.04, rails-2.3.5, ruby 1.8.6. Please help me to remove this little issue. Thanks!</p>			
Related issues:			
Related to Redmine - Defect #10134: Case insensitive search is not working wi...		Confirmed	
Blocked by Redmine - Feature #4050: Ruby 1.9 support		Closed	2009-10-18

History

#1 - 2011-05-27 16:26 - Etienne Massip

- Target version set to Candidate for next minor release

#2 - 2011-05-28 04:28 - Alexey Ivlev

Thank you very much!

#3 - 2011-05-28 09:29 - Etienne Massip

- Target version deleted (Candidate for next minor release)

Sorry but the underlying issue seems to be a Ruby Regexp one as Redmine code in SearchHelper#highlight_tokens seems fairly safe in the way it handles case : [source:trunk/app/helpers/search_helper.rb#L22](https://source.trunk/app/helpers/search_helper.rb#L22).

Added [#4050](#) as blocker.

#4 - 2011-05-28 14:51 - Alexey Ivlev

In other words, the problem will be solved only when the Ruby Regexp will be fixed?

#5 - 2011-05-30 00:33 - Etienne Massip

That's what I think, yes.

#6 - 2011-11-03 16:39 - Yuriy Sokolov

- % Done changed from 0 to 50

Actually, I made a fix

```

module SearchHelper
  def highlight_tokens(text, tokens)
    return text unless text && tokens && !tokens.empty?
    re_tokens = tokens.collect {|t| Regexp.escape(t.mb_chars.downcase)}
    regexp = Regexp.new "(#{re_tokens.join('|')})"
    result = ''
    position = 0
    text = text.mb_chars
    text.downcase.split(regexp).each_with_index do |words, i|
      if result.length > 1200
        # maximum length of the preview reached
        result << '...'
        break
      end
      words = text[position ... (position + words.size)]
      position += words.size
      if i.even?
        result << h(words.length > 100 ? "#{words.slice(0..44)} ... #{words.slice(-45..-1)}" : words)
      else
        t = (tokens.index(words.downcase) || 0) % 4
        result << content_tag('span', h(words), :class => "highlight token-#{t}")
      end
    end
    result
  end
end
end

```

#7 - 2013-10-14 11:55 - Jean-Philippe Lang

- Subject changed from *Highlighting in search results is case sensitive for cyrillic pattern* to *Highlighting in search results is case sensitive for cyrillic pattern*

Files

with_highlighting.png	550 Bytes	2010-10-01	Alexey Ivlev
without_highlighting.png	623 Bytes	2010-10-01	Alexey Ivlev