

Ruby trunk - Feature #15771

Add `String#split` option to set split_type string when a single space separator

04/16/2019 04:42 AM - 284km (kazuma furuhashi)

Status:	Open								
Priority:	Normal								
Assignee:									
Target version:									
Description									
In String#split, when separator is a single space character, it execute as split_type: awk.									
For example, CSV library handles it like this. https://github.com/ruby/csv/blob/7ff57a50e81c368029fa9b664700bec4a456b81b/lib/csv/parser.rb#L508-L512									
<pre>if @column_separator == " ".encode(@encoding) @split_column_separator = Regexp.new(@escaped_column_separator) else @split_column_separator = @column_separator end</pre>									
Unfortunately, in this case regexp is slower than string. For example, the following result is about 9 times slower. https://github.com/284km/benchmarks_no_yatu#stringsplitstring-or-regexp									
<pre>\$ be benchmark-driver string_split_string-regexp.yml --rbenv '2.6.2'</pre> Comparison: <table><tr><td>string:</td><td>3161117.6</td><td>i/s</td><td></td></tr><tr><td>regexp:</td><td>344448.0</td><td>i/s</td><td>- 9.18x slower</td></tr></table>		string:	3161117.6	i/s		regexp:	344448.0	i/s	- 9.18x slower
string:	3161117.6	i/s							
regexp:	344448.0	i/s	- 9.18x slower						
So I want to add the :literal option to run as split_type: string.									
Implementation									
<ul style="list-style-type: none">https://github.com/284km/ruby/tree/split_space<ul style="list-style-type: none">test code: https://github.com/284km/ruby/blob/split_space/test/ruby/test_string.rb#L1708-L1713									
This change will result in the following:									
<pre>" a b c ".split(" ") => ["a", "b", "c"] " a b c ".split(" ", -1) => ["a", "b", "c", ""] " a b c ".split(" ", literal: true) => ["", "a", "", "b", "", "", "c"] " a b c ".split(" ", -1, literal: true) => ["", "a", "", "b", "", "", "c", "", "", "", ""]</pre>									

History

#1 - 04/18/2019 11:15 PM - 284km (kazuma furuhashi)

pull request: <https://github.com/ruby/ruby/pull/2132>