

Re: regular expression question

Source:

<http://www.tech-archive.net/Archive/DotNet/microsoft.public.dotnet.general/2006-03/msg01031.html>

- *From:* "Kevin Spencer" <kevin@xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx>
 - *Date:* Sat, 25 Mar 2006 09:46:08 -0500
-

Hi Ludwig,

It is not possible to answer your question as you've stated it. Here's why:

I'm using the regular expression `\b\w` to find the beginning of a word, in my C# application. If the word is 'public', for example, it works. However, if the word is '<public', it does not work: it seems that '<' is not a valid character, so the beginning of the word starts at the letter 'p' instead of '<'.

You have not defined your terms. You use the word "word," but you have not defined what that is supposed to mean in your situation. In regular expressions, there are no words, only characters. The `"\w"` character class indicates a word `*character*`. A word character is defined in regular expressions as a character that is either a digit or a letter of the alphabet.

So, the character '<' is not defined in regular expressions as a word character, and therefore is not identified as belonging to the set defined by your rule.

However, while you have stated that you **do** want to identify the character '<' as the "beginning of a word," you have not stated exactly what the rule is, only a small part of it. For example, by what you've told me, the following character sequences could all be "words" –

Hello Ludwig ('H', 'L') The first letters of each word are identified.

Hello, <Ludwig> ('H', '<') The first letter of "Hello" and the beginning '<' are identified.

Hello, !!!!! ('H', '!') The first letter of "Hello" and the beginning '!' are identified. This is possible because you have not stated what characters you do **not** consider to be the beginnings of words.

And so on. In other words, a regular expression is shorthand for a rule that

Re: regular expression question

defines a pattern. You need to explicitly define what the rule is in order for me to create a regular expression that satisfies that rule.

--

HTH,

Kevin Spencer
Microsoft MVP
Professional Numbskull

Show me your certification without works,
and I'll show my certification
by my works.

"Ludwig" <none@xxxxxxx> wrote in message
news:v6ea22ttuqisnt52vuvdg0voopljbm8ht@xxxxxxxxxxx

Hi,

i'm using the regular expression `\b\w` to find the beginning of a word, in my C# application. If the word is 'public', for example, it works. However, if the word is '<public', it does not work: it seems that '<' is not a valid character, so the beginning of the word starts at the letter 'p' instead of '<'.

Because I'm not an expert in regular expressions, maybe someone of you guys can help me? I need the correct regex to find the beginning of the word '<public' in a string.

Thanks...

Kind regards,
Ludwig

.