Will our program be presented double quoted phrases inside double quoted phrases as input -- example: "University "of" Alabama"? Also will there ever be a case of a word inside of a parenthesis starting with a double quote-- example: "University "of Alabama". If so, how should we interpret it?

The scanner's readString function takes care of embedded quotes, but embedded quotes need to be escaped with a backslash. So:

"University "of" Alabama"

would be a legal phrase. I'm not sure what you're asking in your second questions, but all quotes will be matched. Also, for all quoted phrases, the double quotes will be the first and last non-whitespace character.

So, just for clarification, you would want this example to be put in as university "of" alabama or as university of alabama?

Embedded quotes should be removed, as are other non-letter (and non-space) characters.

When we are reading commands and words/phrases to be deleted from the text file, will we need to format the word before we search for it in the tree? For example, suppose we have already inserted the word "alabama" into the tree, will we see a command like this: d Ala%bama ,and be
expected to delete the "alabama" key?

Subject: Re: Legal Input  
Posted by lusth on Wed, 05 Oct 2016 23:00:23 GMT  
View Forum Message ◄ Reply to Message

yes

Subject: Re: Legal Input  
Posted by nebrinda on Tue, 11 Oct 2016 16:15:04 GMT  
View Forum Message ◄ Reply to Message

Are unicode characters considered legal input?

Subject: Re: Legal Input  
Posted by lusth on Tue, 11 Oct 2016 20:20:30 GMT  
View Forum Message ◄ Reply to Message

No, the spec says only ASCII letters are preserved.

Subject: Re: Legal Input  
Posted by btlindow on Wed, 12 Oct 2016 00:18:19 GMT  
View Forum Message ◄ Reply to Message

From talking with the ranks prior to class, there is still some disagreements with how to handle escaped punctuation (quotations) within quotes. If the user inputs:

"University "of" Alabama"

Is the expected "trimmed" value of that string to be:

university "of" alabama

or

university of alabama

I suppose it comes down to defining what makes a quote an embedded quote. Sorry for beating a dead horse... and then dragging it around... I just thought I would try to get some black and white clarification for everyone.
I answered this in message #6112 above. Embedded quotes are removed.

Because phrases are to begin and end with double quotes, is it possible that we see something like this, "University of Alabama"!, where there is a character immediately after the last double quote with no whitespace in between?

The scanner would be able to handle that. The readString function would stop at the terminating double quote and readToken would pick up the ! as a token. If you use readString and readToken correctly, everything will work out.

Another Legal Input question, if the string has a trailing space, do we remove it? For example "string " has a space after the g, would this be legal?

The space would be preserved.

Will we ever be presented with a phrase that has a space immediately after the opening quotes
and before the next word? For example, "university of alabama". The instructions state to replace spans of whitespace with a single space, so this would result in a string that has a whitespace as the first character. Would this be the correct way that the string should be formatted?

Subject: Re: Legal Input
Posted by lusth on Fri, 21 Oct 2016 19:39:45 GMT
View Forum Message <> Reply to Message

Whitespace is preserved, but collapsed to a single space.

Subject: Re: Legal Input
Posted by jgmurphy1 on Mon, 24 Oct 2016 03:55:02 GMT
View Forum Message <> Reply to Message

In regards to preserving whitespace, do we retain only as single space in between words of a phrase as "one two" would be "one two" or do we also preserve a single space in front and back of the phrase such as " one two " would be " one two " ?

Subject: Re: Legal Input
Posted by lusth on Mon, 24 Oct 2016 13:25:50 GMT
View Forum Message <> Reply to Message

You always preserve whitespace, consolidating it to a single space.

Subject: Re: Legal Input
Posted by jgmurphy1 on Mon, 24 Oct 2016 18:23:40 GMT
View Forum Message <> Reply to Message

I believe the confusion of preserving whitespace is coming from the example command calls in the project description. It says that

i "spongebob squarepants"

is equivalent to

i "spongebob squarepants"

which shows that you do remove any trailing whitespace

a snippet of the project details I am referring to is attached
Can null strings be passed in as input? (double quotes with nothing between, "")

No. I won't test that kind of input.

should the command letters (i,s,r,f,d) be inserted? For example, a command file that reads:{ i i } would insert i into the tree?

Thanks

I believe it should. Since you should only check for the instruction and whatever after should be valid input. I could easily have assumed wrong though.

If there is an input that consists of all non-alpha or whitespace characters (ex. "@#$"), it will be cleaned to an empty string. Should we check for this case and move on when it happens, or will we not be presented with that input.

This is an extension of sestephens question above.
If there are any whitespace characters in a phrase the cleaned version will have at least one space in it.
If a phrase is nothing but non-alpha character, then the resulting cleaned string will have a length of zero, and the specs say to "not insert empty strings into the tree".

Here is my interpretation of whitespace input... It is essentially a reference to the answers in this thread:

1231232 (or any string of numbers) will not be inserted into the tree
"1231232" will not be inserted into the tree

"1232 1232" will result in " " being inserted into the tree

"" will not be checked
" " will be inserted into the tree as " " (a single white space)
Alternatively, any amount of whitespace in between double quotes will be inserted into the tree as " "

" phrase of words " will be inserted as shown, but if more than one whitespace is before, in between, or after an word, it will removed and end up looking like " phrase of words "

"University "of" Alabama" will be inserted into the tree as "university of alabama"
if "University "of" Alabama" is in the corpus file, the scanner module returns an error and exits

Please let me know if you see anything wrong with this post... I did it to consolidate this thread for future readers and to help myself remember everything.

That last example you have is legal input I believe.
If you are using readToken and readString properly, that last two examples should work. I won't intentionally test for that, though.

Subject: Re: Legal Input  
Posted by berryhilleric on Tue, 01 Nov 2016 00:44:26 GMT  
View Forum Message <> Reply to Message

Is it not possible to have a string end up with several whitespace characters, for example: " u of alabama " would actually contain 4 whitespace characters according to the spec?

Subject: Re: Legal Input  
Posted by sestephens on Tue, 01 Nov 2016 05:26:50 GMT  
View Forum Message <> Reply to Message

Quote:
"University "of" Alabama" will be inserted into the tree as "university of alabama" if "University "of" Alabama" is in the corpus file, the scanner module returns an error and exits

"University "of" Alabama" and "University "of" Alabama" would both be inserted as "university " then "of" then " alabama"

Quote: Is it not possible to have a string end up with several whitespace characters, for example: " u of alabama " would actually contain 4 whitespace characters according to the spec?

Not 100% sure I understand your question, but " u of alabama " would be inserted as exactly that, with 4 spaces.

Subject: Re: Legal Input  
Posted by jgmills on Tue, 01 Nov 2016 20:25:11 GMT  
View Forum Message <> Reply to Message

Dr. Lusth mentioned something about having new lines (\n) and tabs (\t) when reading in strings. How should those be handled? Are they supposed to be condensed into a single whitespace? Or removed altogether just like a Non-ASCII letter?

Subject: Re: Legal Input  
Posted by jgmurphy1 on Tue, 01 Nov 2016 21:40:33 GMT  
View Forum Message <> Reply to Message

They are considered whitespace and should be condensed accordingly.