Bug in Matlab implementation of Porter Stemmer

I use the Matlab implementation of the Porter Stemming algorithm published here. When I use it on a large portion of text, I get an error message when the string `"ies"` is passed in.

```
>> porterStemmer('ies')
```

I get the error message:

```
Subscript indices must either be real positive integers or
logicals.

Error in porterStemmer>step2 (line 274)
switch b(k-1)

Error in porterStemmer (line 42)
    x = step2(x{1}, x{2}, k0);
```

I suppose it might be, because `"ies"` is a common word ending and also appears in the script on line 223 as `elseif ends('ies', b, k)`.

I was expecting the program to be robust to cases like this. **Was I wrong to think that or this is a bug?**

For the time being, I've resorted to using the Porter Stemmer 2 (here), which doesn't throw an error and stems "ies" to "ie". Stemming "ies" is a bit pointless obviously, but I'd rather not have it throw errors in edge cases like this.

matlab   nlp   text-mining   information-retrieval   porter-stemmer

asked Dec 14 '15 at 19:14

ulima2_
347   1   7

You should debug your code and see what causes the error. Either put a breakpoint on line 274 in `porterStemmer`, or just execute this code `dbstop if error` and inspect the variables when the debugger encounters an error. The error means that someone tried indexing an array using a non-natural index (complex/imaginary, non-integer, or 0). – Dev-iL Dec 14 '15 at 19:30

If you do verify it's a bug, be sure to notify the owner of the code. – Carl Witthoft Dec 15 '15 at 14:43