

# 1.3 RAT

1. We see that the words 'd' and 's' appear frequently in word tokenization results. Why is this?
2. What is the output of:  

```
>>> '_' .join("this\tlittle\tpiggy".split('\t'))
```
3. The Porter stemmer would stem the word 'assembling' as  
A) assembles                      C) assembling  
B) assemble                        D) assembl
4. A period (".") occurs at the end of four words that each have one of the following four wordshapes. Which of the four periods (all else being equal) is more likely than the other to be a sentence boundary?  
A) Upper                            C) Cap  
B) Number                          D) Lower