

# Parts of Speech Assessment

For this assessment we'll be using the short story [The Tale of Peter Rabbit](#) by Beatrix Potter (1902). The story is in the public domain; the text file was obtained from [Project Gutenberg](#).

```
In [1]: # RUN THIS CELL to perform standard imports:
import spacy
nlp = spacy.load('en_core_web_sm')
from spacy import displacy
```

1. Create a Doc object from the file `peterrabbit.txt`

HINT: Use `with open('../TextFiles/peterrabbit.txt') as f:`

In [2]:

2. For every token in the third sentence, print the token text, the POS tag, the fine-grained TAG tag, and the description of the fine-grained tag.

In [3]: *# Enter your code here:*

They	PRON	PRP	pronoun, personal
lived	VERB	VBD	verb, past tense
with	ADP	IN	conjunction, subordinating or preposition
their	ADJ	PRP\$	pronoun, possessive
Mother	PROPN	NNP	noun, proper singular
in	ADP	IN	conjunction, subordinating or preposition
a	DET	DT	determiner
sand	NOUN	NN	noun, singular or mass
-	PUNCT	HYPH	punctuation mark, hyphen
bank	NOUN	NN	noun, singular or mass
,	PUNCT	,	punctuation mark, comma
underneath	ADP	IN	conjunction, subordinating or preposition
the	DET	DT	determiner
root	NOUN	NN	noun, singular or mass
of	ADP	IN	conjunction, subordinating or preposition
a	DET	DT	determiner

	SPACE		None
very	ADV	RB	adverb
big	ADJ	JJ	adjective
fir	NOUN	NN	noun, singular or mass
-	PUNCT	HYPH	punctuation mark, hyphen
tree	NOUN	NN	noun, singular or mass
.	PUNCT	.	punctuation mark, sentence closer

SPACE \_SP None

3. Provide a frequency list of POS tags from the entire document

In [4]:

```
83. ADJ : 83
84. ADP : 127
85. ADV : 75
88. CCONJ: 61
89. DET : 90
91. NOUN : 176
92. NUM : 8
93. PART : 36
94. PRON : 72
95. PROPN: 75
96. PUNCT: 174
99. VERB : 182
102. SPACE: 99
```

4. CHALLENGE: What percentage of tokens are nouns?

HINT: the attribute ID for 'NOUN' is 91

In [5]:

```
176/1258 = 13.99%
```

5. Display the Dependency Parse for the third sentence

In [6]:

```
They PRON lived VERB with ADP their ADJ Mother PROPN in ADP a DET sand- NOUN bank, NOUN underneath ADP the DET root NOUN of ADP a DET
very ADV big ADJ fir- NOUN tree. NOUN SPACE nsubj prep poss pobj prep det compound pobj prep det pobj prep det advmod amod compound punct
```

\*6. Show the first two named entities from Beatrix Potter's *The Tale of Peter Rabbit* \*

In [7]:

```
The Tale of Peter Rabbit - WORK_OF_ART - Titles of books, songs, etc.
Beatrix Potter - PERSON - People, including fictional
```

7. How many sentences are contained in *The Tale of Peter Rabbit*?

In [8]:

Out[8]: 56

8. CHALLENGE: How many sentences contain named entities?

In [9]:

Out[9]: 49

9. CHALLENGE: Display the named entity visualization for `list_of_sents[0]` from the previous problem

In [10]:

```
The Tale of Peter Rabbit WORK_OF_ART , by Beatrix Potter PERSON ( 1902 DATE ).
```