

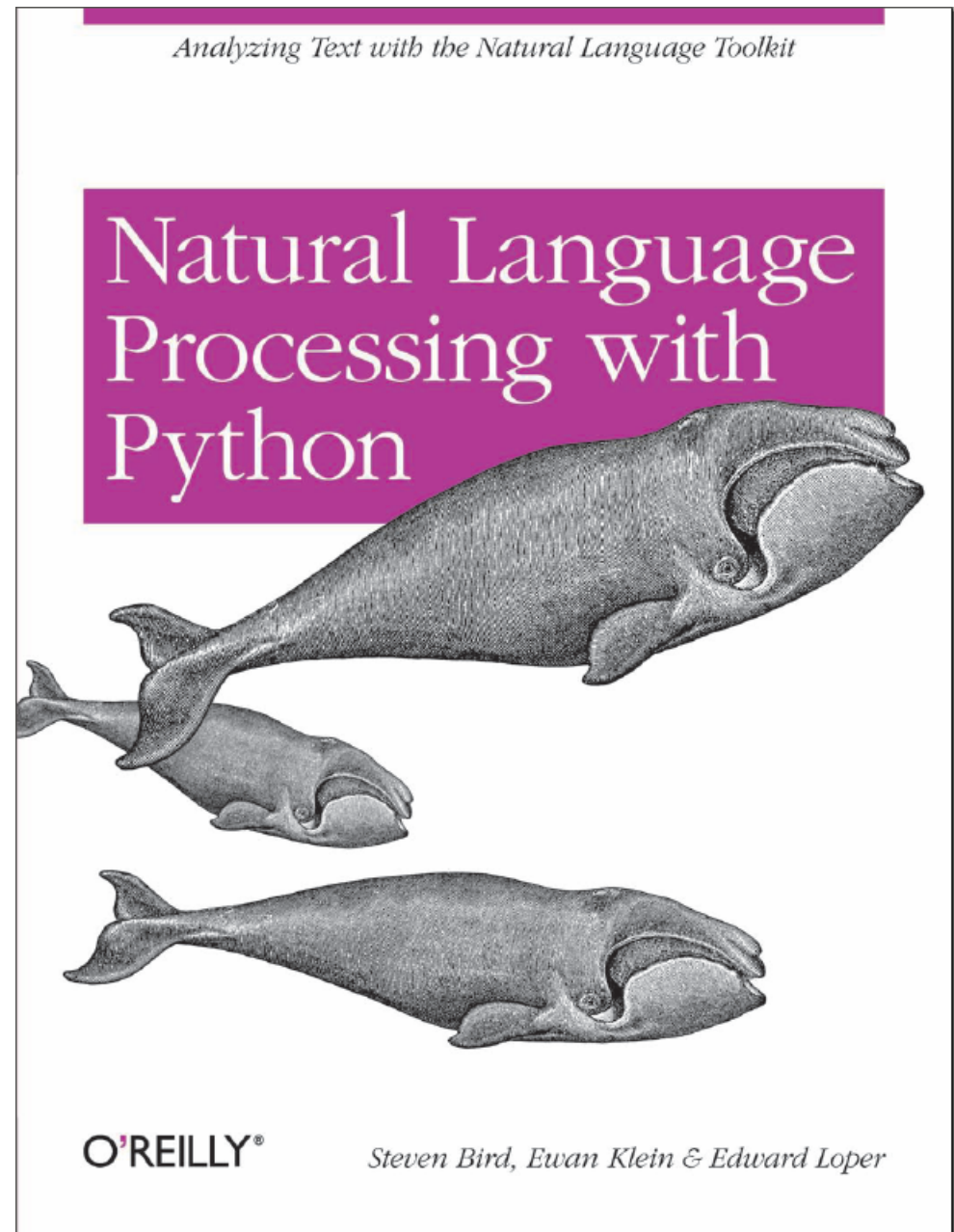
**Steven Bird, Ewan Klein,  
and Edward Loper**

Copyright © 2009. All rights reserved.

Printed in the United States of America.

Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

<http://code.google.com/p/nltk/>



# NLTK 正则表达式应用例子

```
>>> import nltk
```

```
>>> import re
```

```
>>> wsj = sorted(set(nltk.corpus.treebank.words()))
```

```
>>> [w for w in wsj if re.search('^[0-9]+\.[0-9]+$', w)]
```

```
['0.0085', '0.05', '0.1', '0.16', '0.2', '0.25', '0.28', '0.3',  
'0.4', '0.5', '0.50', '0.54', '0.56', '0.60', '0.7', '0.82',  
'0.84', '0.9', '0.95', '0.99', '1.01', '1.1', '1.125', '1.14',  
'1.1650', '1.17', '1.18', '1.19', '1.2', ...]
```

# NLTK 正则表达式应用例子（续）

```
>>> [w for w in wsj if re.search('^[0-9]{4}$', w)]
```

```
['1614', '1637', '1787', '1901', '1903', '1917', '1925', '1929',  
'1933', ...]
```

```
>>> [w for w in wsj if re.search('(ed|ing)$', w)]
```

```
['62%-owned', 'Absorbed', 'According', 'Adopting',  
'Advanced', 'Advancing', ...]
```