

```
# load text
filename = 'novel.txt'
file = open(filename, 'rt')
text = file.read()
print(text)
file.close()

words = text.split()
print(words[:50])

import re

filename = 'novel.txt'
file = open(filename, 'rt')
text = file.read()
file.close()

words = re.split(r'\W+', text)
print(words[:50])

import string
print(string.punctuation)

stripped = [re.sub(r'^\w\s', '', w) for w in words]
print(stripped[:50])

words = [word.lower() for word in stripped]
print(words[:50])
```