

**BIS 420 PROGRAMMING FOR DATA SCIENCE**  
**PRAJAKTA POHARE**  
**CHAPTER 13 EXERCISE 13.4**  
**ILLINOIS STATE UNIVERSITY**

Modify the previous program to read a word list (see Section 9.1) and then print all the words in the book that are not in the word list. How many of them are typos? How many of them are common words that should be in the word list, and how many of them are really obscure?

```
/usr/local/bin/python3 "/Users/prajaktapohare/Library/CloudStorage/OneDrive-ILStateUniversity/BIS420/Week 13/BIS420_PrajaktaPohare_Ch13_13.4.py"
~ /usr/local/bin/python3 "/Users/prajaktapohare/Library/CloudStorage/OneDrive-ILStateUniversity/BIS420/Week 13/BIS420_PrajaktaPohare_Ch13_13.4.py"

Analysis of book.txt:
Total words: 4560
Unique words: 1091
Unknown words (not in dictionary): 67
  Typos (short/invalid): 40
  Common but missing: 0
  Obscure/rare words: 27

Sample typos: ['1', '10', '11', '12', '13', '14', '1435', '15', '16', '17']
Sample common missing: []
Sample obscure words: ['100b0', '93i', '93old', '93rhythmically94', '93rhythmically94', '95b0', 'administered', 'age94', 'america', 'benson']
~
```