

```
1: #!/usr/bin/env python
2: """reducer.py"""
3:
4: from operator import itemgetter
5: import sys
6:
7: current_word = None
8: current_count = 0
9: word = None
10:
11: # input comes from STDIN
12: for line in sys.stdin:
13:     # remove leading and trailing whitespace
14:     line = line.strip()
15:
16:     # parse the input we got from mapper.py
17:     word, count = line.split('\t', 1)
18:
19:     # convert count (currently a string) to int
20:     try:
21:         count = int(count)
22:     except ValueError:
23:         # count was not a number, so silently
24:         # ignore/discard this line
25:         continue
26:
27:     # this IF-switch only works because Hadoop sorts map output
28:     # by key (here: word) before it is passed to the reducer
29:     if current_word == word:
30:         current_count += count
31:     else:
32:         if current_word:
33:             # write result to STDOUT
34:             print '%s\t%s' % (current_word, current_count)
35:         current_count = count
36:         current_word = word
37:
38: # do not forget to output the last word if needed!
39: if current_word == word:
40:     print '%s\t%s' % (current_word, current_count)
```