MAPPER :

```
#!/usr/bin/python
"""mapper.py"""

import sys
for line in sys.stdin:
        line = line.strip()
        words = line.split()
        for word in words:
                print '%s\t%s' % (word, 1)
```

REDUCER:

```
#!/usr/bin/env python
"""reducer.py"""

import sys

current_word = None
current_count = 0
word = None

# input comes from STDIN
for line in sys.stdin:
   # remove leading and trailing whitespace
   line = line.strip()

   # parse the input we got from mapper.py
   word, count = line.split('\t', 1)

   # convert count (currently a string) to int
   try:
```

```
        count = int(count)
    except ValueError:
        # count was not a number, so silently
        # ignore/discard this line
        continue

    # this IF-switch only works because Hadoop sorts map output
    # by key (here: word) before it is passed to the reducer
    if current_word == word:
        current_count += count
    else:
        if current_word:
            # write result to STDOUT
            print '%s\t%s' % (current_word, current_count)
        current_count = count
        current_word = word

# do not forget to output the last word if needed!
if current_word == word:
    print '%s\t%s' % (current_word, current_count)
```

Terminal Commands:

```
gedit mapper.py
gedit word.txt
gedit reducer.py
cat word.txt | python mapper.py | sort | python reducer.py
```