LABWORK 4: WordCount

Student's name: Nguyen Trung Kien

Student's ID: BI12-224

1. System architecture

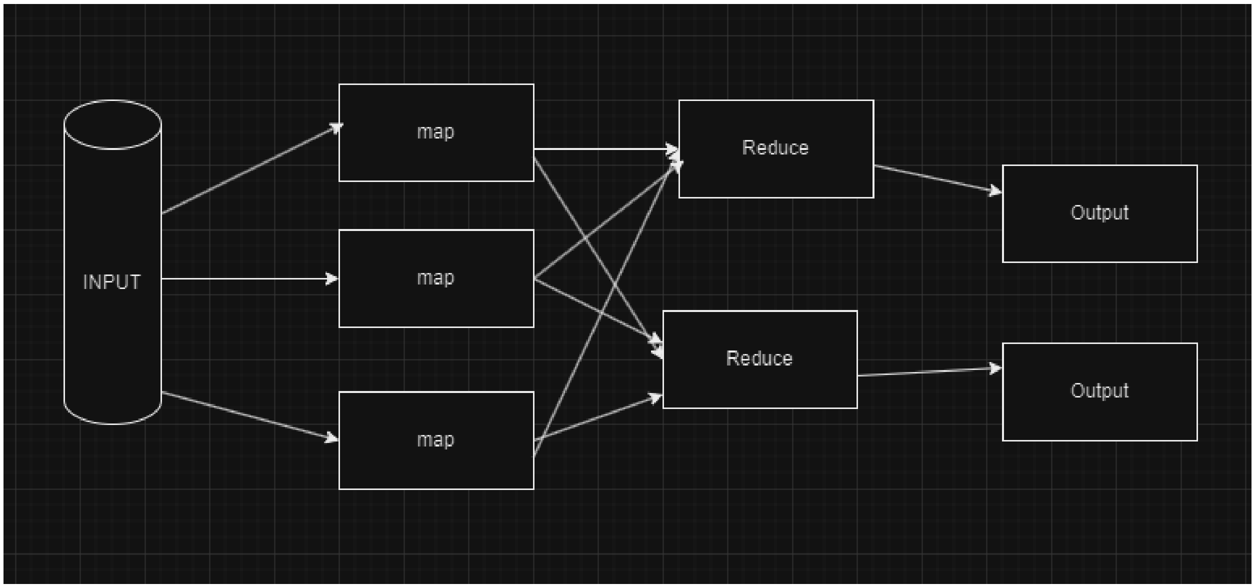
a. Code explain

- Map Function (map):
    i. This function tokenizes a line of text into words and creates a key-value pair for each word encountered.
    ii. It takes a line of text line, an array of KeyValue pairs kv, and a pointer to the count of key-value pairs kv_count.
    iii. It tokenizes the line using strtok function with common delimiters (space, period, comma, tab, newline, carriage return).
    iv. For each token (word), it copies the word into the key of the key-value pair and sets its count to 1.
    v. It increments the count of key-value pairs.
- Reduce Function (reduce):
    i. This function aggregates the key-value pairs to calculate the total count for each unique word.
    ii. It takes an array of KeyValue pairs input and its size input_size.
    iii. It initializes a new array result to store the aggregated key-value pairs and sets its size to 0.
    iv. It iterates through each key-value pair in the input array.
    v. For each pair, it checks if the word already exists in the result array. If it does, it increments its count. Otherwise, it adds a new key-value pair to the result array.
    vi. Finally, it prints each word along with its count.
- Main Function:

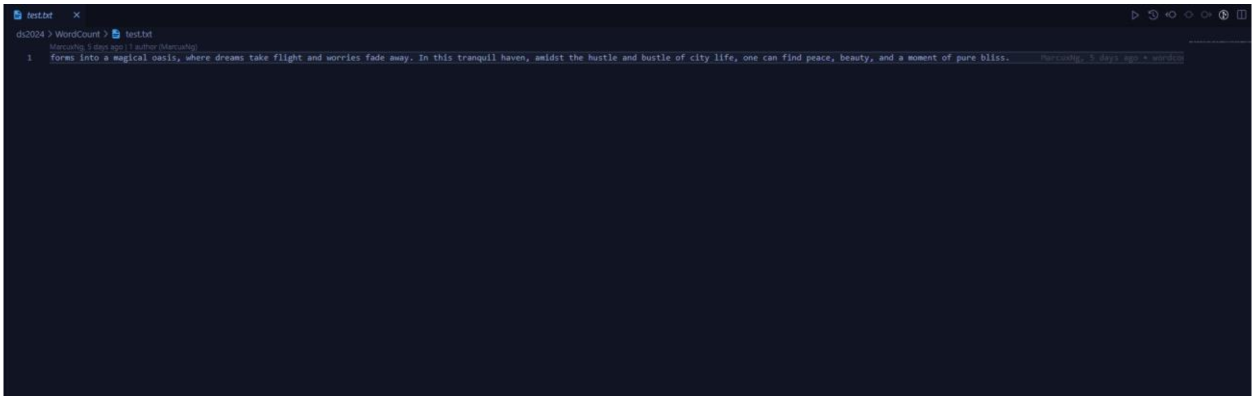
i. It declares variables for file handling, reading lines, storing key-value pairs, and counting key-value pairs.
ii. It opens the file "test.txt" for reading.
iii. It reads each line from the file and calls the map function to tokenize the line and create key-value pairs.
iv. After reading all lines, it closes the file.
v. It then calls the reduce function to aggregate the key-value pairs and print the word count.

b. Workflow

- The main function reads each line from the file, tokenizes it into words, and creates key-value pairs.
- The reduce function aggregates the key-value pairs to count the occurrences of each word.
- Finally, the program prints the word count.

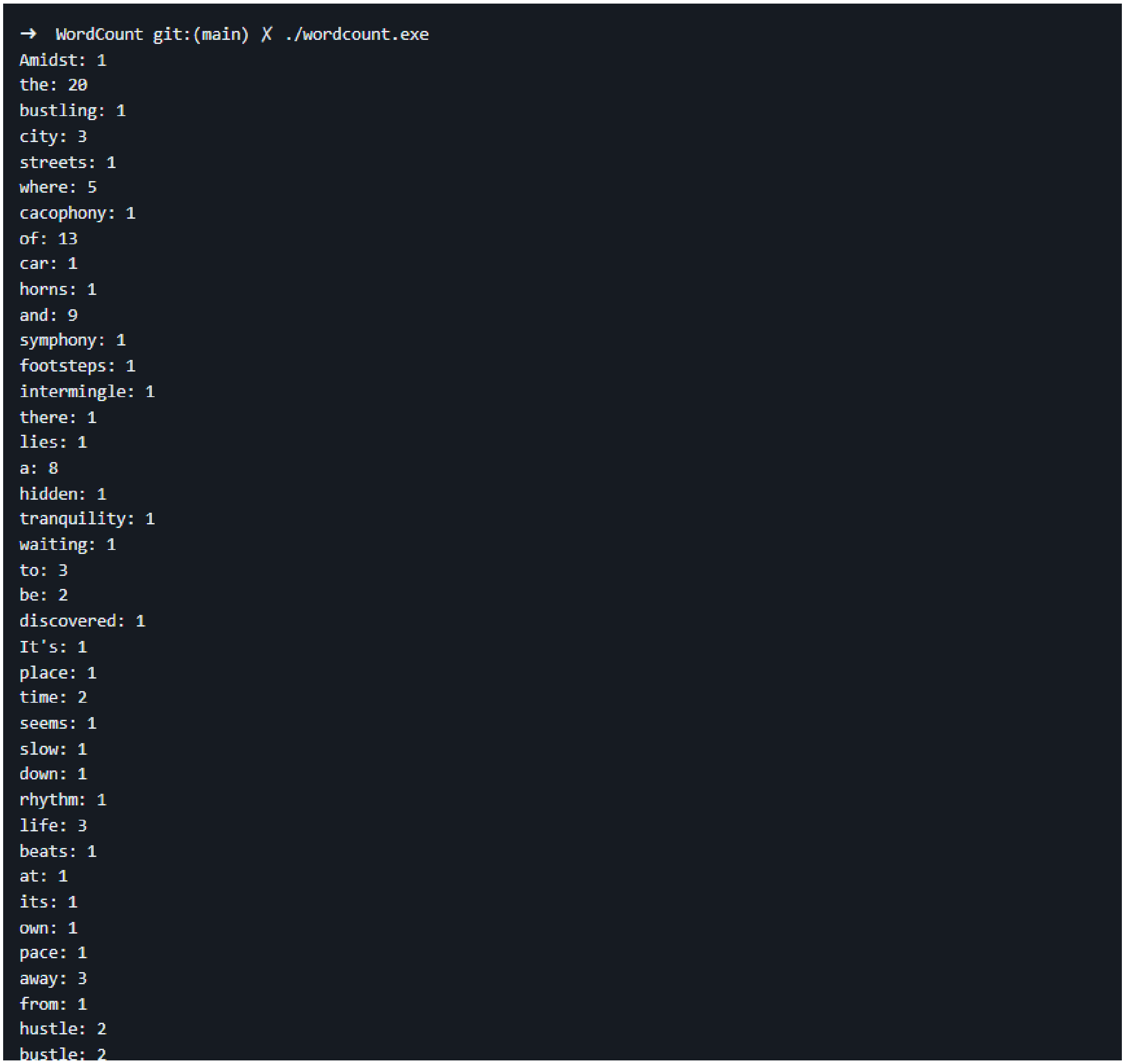


2. Implementation

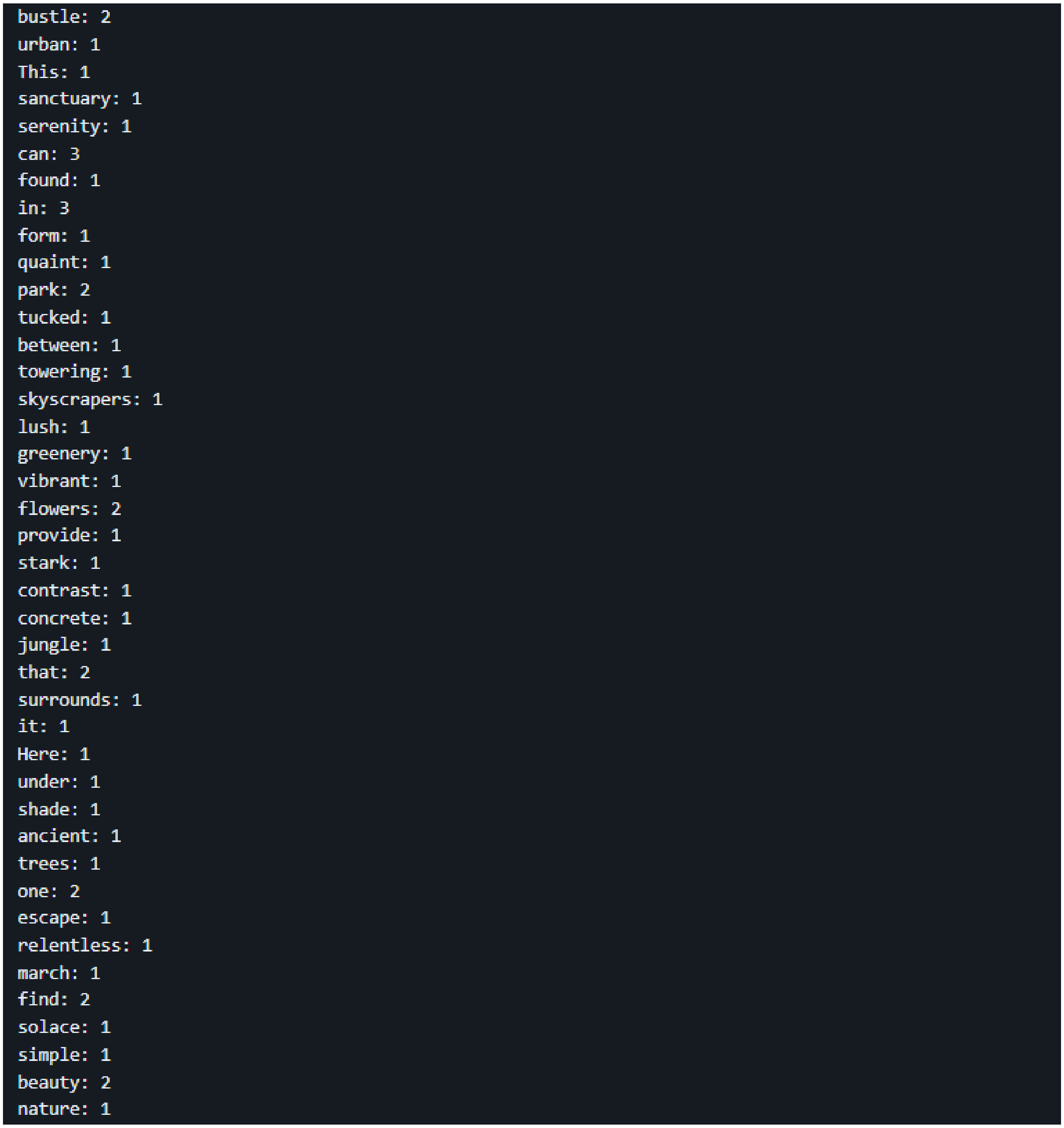

test.txt
ds2024 > WordCount > test.txt
MarcuxNg, 5 days ago | 1 author (MarcuxNg)
MarcuxNg, 5 days ago • wordco

```
forms into a magical oasis, where dreams take flight and worries fade away. In this tranquil haven, amidst the hustle and bustle of city life, one can find peace, beauty, and a moment of pure bliss.
```

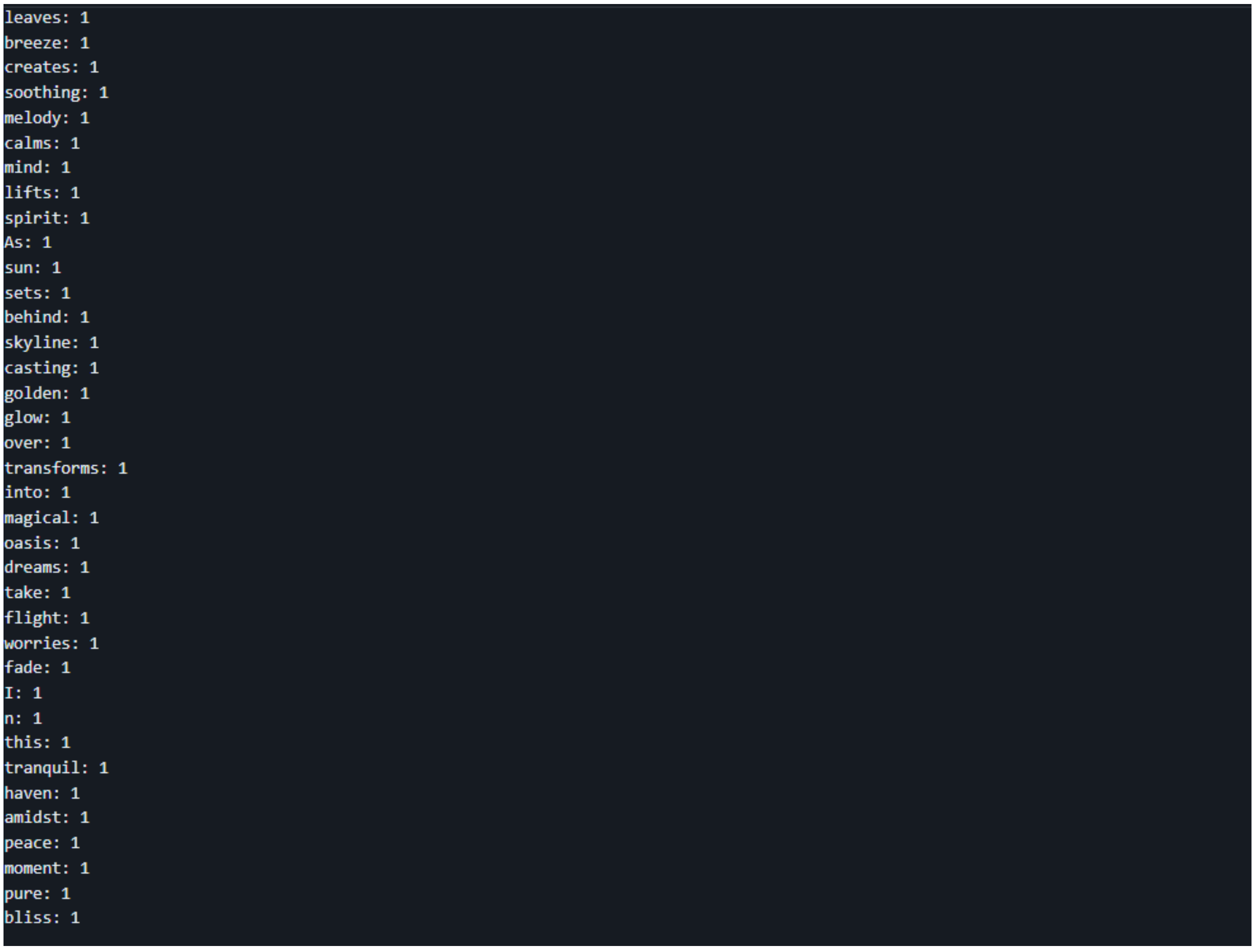
Figure 1. test.txt



```
➜  WordCount git:(main) ✗ ./wordcount.exe
Amidst: 1
the: 20
bustling: 1
city: 3
streets: 1
where: 5
cacophony: 1
of: 13
car: 1
horns: 1
and: 9
symphony: 1
footsteps: 1
intermingle: 1
there: 1
lies: 1
a: 8
hidden: 1
tranquility: 1
waiting: 1
to: 3
be: 2
discovered: 1
It's: 1
place: 1
time: 2
seems: 1
slow: 1
down: 1
rhythm: 1
life: 3
beats: 1
at: 1
its: 1
own: 1
pace: 1
away: 3
from: 1
hustle: 2
bustle: 2
```



```
bustle: 2
urban: 1
This: 1
sanctuary: 1
serenity: 1
can: 3
found: 1
in: 3
form: 1
quaint: 1
park: 2
tucked: 1
between: 1
towering: 1
skyscrapers: 1
lush: 1
greenery: 1
vibrant: 1
flowers: 2
provide: 1
stark: 1
contrast: 1
concrete: 1
jungle: 1
that: 2
surrounds: 1
it: 1
Here: 1
under: 1
shade: 1
ancient: 1
trees: 1
one: 2
escape: 1
relentless: 1
march: 1
find: 2
solace: 1
simple: 1
beauty: 2
nature: 1
```



```
leaves: 1
breeze: 1
creates: 1
soothing: 1
melody: 1
calms: 1
mind: 1
lifts: 1
spirit: 1
As: 1
sun: 1
sets: 1
behind: 1
skyline: 1
casting: 1
golden: 1
glow: 1
over: 1
transforms: 1
into: 1
magical: 1
oasis: 1
dreams: 1
take: 1
flight: 1
worries: 1
fade: 1
I: 1
n: 1
this: 1
tranquil: 1
haven: 1
amidst: 1
peace: 1
moment: 1
pure: 1
bliss: 1
```