ASSIGNMENT-2 : MAP-REDUCE

In [23]:
```
# @title
from collections import defaultdict
import re

# Map function to emit word occurrences
def map_words(file_content):
    word_counts = defaultdict(int)
    words = re.findall(r'\w+', file_content.lower())
    for word in words:
        word_counts[word] += 1
    return word_counts

# Reduce function to aggregate the word counts
def reduce_word_counts(mapped_word_counts):
    reduced_word_counts = defaultdict(int)
    for word, count in mapped_word_counts.items():
        reduced_word_counts[word] += count
    return reduced_word_counts

# Read the file and apply MapReduce
with open('/file1 (1).txt', 'r', encoding='ISO-8859-1') as file:
    file_content = file.read()

# Map phase
mapped_words = map_words(file_content)

# Reduce phase
word_counts = reduce_word_counts(mapped_words)

# Display word counts
for word, count in word_counts.items():
    print(f"{word}: {count}")
```

```
harry: 11
potter: 11
and: 73
the: 170
half: 10
blood: 10
prince: 10
â: 225
j: 10
k: 10
rowling: 10
to: 60
him: 12
though: 4
this: 12
had: 57
been: 14
nothing: 3
how: 2
he: 51
felt: 2
when: 8
a: 65
self: 1
proclaimed: 1
wizard: 4
bounced: 1
out: 17
of: 75
fireplace: 3
shaken: 1
his: 43
hand: 5
remained: 2
speechless: 1
throughout: 1
fudgeâ: 3
s: 36
kindly: 1
explanation: 1
that: 43
there: 7
were: 6
witches: 1
wizards: 2
still: 6
living: 2
in: 34
secret: 1
all: 17
over: 3
world: 3
reassurances: 1
was: 29
not: 25
bother: 3
head: 3
about: 11
them: 4
as: 11
ministry: 4
magic: 4
took: 3
responsibility: 1
```

```
for: 19
whole: 1
wizarding: 2
community: 2
prevented: 1
non: 2
magical: 6
population: 3
from: 8
getting: 1
wind: 1
it: 25
said: 27
fudge: 43
difficult: 1
job: 1
encompassed: 1
everything: 1
regulations: 1
on: 12
responsible: 1
use: 1
broomsticks: 1
keeping: 1
dragon: 1
under: 3
control: 3
prime: 53
minister: 51
remembered: 1
clutching: 1
desk: 2
support: 2
at: 17
point: 4
then: 9
patted: 1
shoulder: 1
dumbstruck: 1
fatherly: 1
sort: 1
way: 2
worry: 4
itâ: 7
odds: 1
youâ: 6
ll: 4
never: 6
see: 3
me: 8
again: 6
iâ: 8
only: 2
you: 25
if: 5
thereâ: 2
something: 3
really: 5
serious: 2
going: 5
our: 3
end: 1
thatâ: 2
likely: 1
```

```
affect: 1
muggles: 4
i: 25
should: 3
say: 8
otherwise: 1
live: 2
let: 2
must: 10
re: 7
taking: 1
lot: 3
better: 2
than: 6
your: 2
predecessor: 1
tried: 3
throw: 1
window: 1
thought: 4
hoax: 2
planned: 1
by: 6
opposition: 2
found: 2
voice: 1
last: 4
desperate: 1
hope: 4
no: 8
gently: 1
m: 5
afraid: 2
look: 2
turned: 1
ministerâ: 5
teacup: 2
into: 9
gerbil: 2
but: 21
breathlessly: 1
watching: 1
chewing: 1
corner: 3
next: 2
speech: 1
why: 5
has: 4
nobody: 1
told: 3
reveals: 1
or: 6
herself: 1
muggle: 4
day: 3
poking: 1
wand: 2
back: 4
inside: 2
jacket: 1
we: 13
find: 3
best: 1
maintain: 1
```

```
secrecy: 1
bleated: 1
hasnâ: 1
t: 19
former: 1
warned: 1
actually: 2
laughed: 1
my: 3
dear: 1
are: 6
ever: 1
tell: 5
anybody: 1
chortling: 1
thrown: 1
some: 1
powder: 1
stepped: 1
emerald: 1
flames: 2
vanished: 2
with: 12
whooshing: 1
sound: 1
stood: 4
quite: 3
motionless: 2
realized: 1
would: 8
long: 1
lived: 2
dare: 1
mention: 1
encounter: 2
soul: 1
who: 22
wide: 1
believe: 1
shock: 1
taken: 2
little: 4
while: 2
wear: 2
off: 2
time: 7
convince: 1
himself: 5
indeed: 1
hallucination: 1
brought: 1
lack: 1
sleep: 1
during: 1
grueling: 1
election: 1
campaign: 1
vain: 1
attempt: 2
rid: 1
reminders: 1
uncomfortable: 1
given: 1
delighted: 1
```

```
niece: 1
instructed: 1
private: 1
secretary: 1
take: 1
down: 7
portrait: 3
ugly: 1
man: 4
announced: 2
arrival: 2
dismay: 1
however: 3
proved: 1
impossible: 1
remove: 1
several: 2
carpenters: 1
builder: 1
two: 3
an: 7
art: 1
historian: 1
chancellor: 1
exchequer: 1
unsuccessfully: 1
prise: 1
wall: 2
abandoned: 1
simply: 2
resolved: 1
thing: 2
silent: 1
rest: 1
term: 1
office: 5
occasionally: 1
could: 5
have: 14
sworn: 1
saw: 2
eye: 1
occupant: 1
painting: 1
yawning: 1
else: 2
scratching: 1
nose: 1
even: 2
once: 4
twice: 1
walking: 2
frame: 1
leaving: 1
stretch: 1
muddy: 1
brown: 1
canvas: 1
behind: 1
trained: 1
picture: 1
very: 4
much: 1
always: 2
```

```
firmly: 1
eyes: 1
playing: 1
tricks: 1
anything: 1
like: 5
happened: 4
three: 5
years: 4
ago: 1
night: 2
tonight: 2
alone: 3
imminent: 1
burst: 1
sopping: 1
wet: 1
state: 1
considerable: 1
panic: 1
before: 5
ask: 2
dripping: 1
axminster: 1
started: 1
ranting: 1
prison: 1
heard: 1
named: 8
seriousâ: 1
black: 7
sounded: 2
hogwarts: 1
boy: 1
called: 2
none: 1
which: 2
made: 2
remotest: 1
sense: 1
ve: 6
just: 5
come: 2
azkaban: 4
panted: 1
tipping: 1
large: 2
amount: 1
water: 1
rim: 1
bowler: 5
hat: 3
pocket: 1
middle: 2
north: 1
sea: 1
know: 12
nasty: 2
flight: 1
dementors: 3
uproarâ: 1
shuddered: 1
theyâ: 1
breakout: 3
```

```
anyway: 4
blackâ: 3
known: 1
killer: 1
may: 2
be: 20
planning: 1
rejoin: 1
course: 3
donâ: 6
is: 11
gazed: 1
hopelessly: 1
moment: 3
well: 6
sit: 3
d: 1
fill: 1
whiskey: 3
rather: 3
resented: 1
being: 2
own: 2
offered: 1
sat: 1
nevertheless: 1
pulled: 1
conjured: 1
glasses: 1
full: 1
amber: 1
liquid: 1
thin: 2
air: 2
pushed: 1
one: 6
drew: 1
up: 10
chair: 1
talked: 1
more: 6
hour: 1
refused: 1
certain: 1
name: 2
aloud: 1
wrote: 1
instead: 1
piece: 1
parchment: 1
thrust: 1
free: 1
leave: 1
too: 2
so: 6
think: 5
squinted: 1
left: 2
lord: 1
vol: 1
snarled: 1
sorry: 2
alive: 5
dumbledore: 2
```

```
says: 1
fastened: 1
pin: 1
striped: 1
cloak: 1
chin: 1
weâ: 8
heâ: 7
dangerous: 2
unless: 2
got: 4
ought: 2
worrying: 1
put: 3
warning: 1
excellent: 1
each: 4
other: 5
good: 3
they: 6
seen: 2
less: 2
year: 1
later: 2
harassed: 1
looking: 2
appeared: 2
cabinet: 1
room: 3
inform: 1
spot: 1
kwidditch: 1
what: 9
cup: 1
involved: 1
fact: 3
whoâ: 1
mark: 1
meant: 1
sure: 1
isolated: 1
incident: 1
liaison: 1
dealing: 1
memory: 2
modifications: 1
spoke: 1
oh: 2
almost: 4
forgot: 1
added: 2
importing: 1
foreign: 1
dragons: 3
sphinx: 2
triwizard: 1
tournament: 1
routine: 1
department: 3
regulation: 2
creatures: 3
tells: 1
rule: 1
book: 1
```

```
notify: 1
bringing: 1
highly: 1
country: 3
spluttered: 1
yes: 4
hoped: 1
beyond: 1
sphinxes: 1
worst: 2
erupted: 1
fire: 2
yet: 2
news: 2
mass: 3
repeated: 2
hoarsely: 1
need: 2
shouted: 1
already: 2
foot: 1
rounded: 1
shout: 1
now: 7
wait: 1
shower: 1
green: 2
sparks: 1
whatever: 1
press: 1
might: 2
foolish: 1
escaped: 1
notice: 1
despite: 1
assurances: 1
their: 2
first: 1
meeting: 1
seeing: 1
nor: 1
becoming: 1
flustered: 1
visit: 1
liked: 1
help: 1
fear: 1
graver: 1
sight: 1
therefore: 1
stepping: 1
disheveled: 1
fretful: 1
sternly: 1
surprised: 1
did: 5
exactly: 2
extremely: 1
gloomy: 1
week: 1
whatâ: 1
er: 3
snapped: 1
run: 1
```

```
enough: 2
concerns: 2
without: 2
same: 1
interrupted: 1
brockdale: 2
bridge: 2
didnâ: 2
wasnâ: 6
hurricane: 3
those: 3
murders: 1
work: 1
herbert: 1
chorleyâ: 1
family: 1
safer: 1
currently: 1
making: 2
arrangements: 1
transferred: 1
st: 1
mungoâ: 1
hospital: 1
maladies: 1
injuries: 2
move: 1
effected: 1
do: 2
blustered: 1
great: 2
deep: 1
breath: 1
am: 1
backâ: 1
mean: 3
groped: 1
details: 2
horrible: 2
conversation: 1
previously: 1
feared: 1
above: 1
others: 1
committed: 2
thousand: 1
terrible: 2
crimes: 1
mysterious: 1
disappearance: 1
fifteen: 1
earlier: 1
canâ: 2
killed: 3
understand: 1
wonâ: 2
explain: 1
properly: 1
certainly: 1
body: 1
talking: 2
killing: 3
suppose: 2
purposes: 1
```

```
discussion: 1
persistent: 1
habit: 1
wishing: 1
appear: 1
informed: 1
any: 5
subject: 1
came: 1
cast: 1
around: 4
remember: 1
previous: 1
conversations: 1
distractedly: 1
turning: 1
rapidly: 2
fingers: 1
sirius: 1
merlinâ: 1
beard: 1
dead: 2
turns: 1
mistaken: 1
innocent: 1
after: 2
league: 1
either: 1
defensively: 1
spinning: 2
faster: 1
evidence: 2
pointed: 1
fifty: 1
eyewitnesses: 1
murdered: 2
matter: 2
premises: 1
inquiry: 1
surprise: 1
fleeting: 1
stab: 1
pity: 1
eclipsed: 1
immediately: 1
glow: 1
smugness: 1
deficient: 1
area: 1
materializing: 1
fireplaces: 1
murder: 2
government: 1
departments: 1
charge: 1
surreptitiously: 1
touched: 1
wood: 1
continued: 1
war: 2
steps: 1
nervously: 1
surely: 1
bit: 1
```

```
overstatement: 1
joined: 1
followers: 3
broke: 1
january: 1
speaking: 1
twirling: 1
fast: 1
lime: 1
blur: 1
since: 1
moved: 1
open: 1
wreaking: 1
havoc: 1
threatened: 1
aside: 1
grief: 1
fault: 3
people: 2
having: 1
answer: 1
questions: 1
rusted: 1
rigging: 1
corroded: 1
expansion: 1
joints: 1
furiously: 2
coloring: 1
saying: 1
caved: 1
blackmail: 1
maybe: 2
standing: 1
striding: 1
efforts: 1
catching: 2
blackmailer: 1
such: 1
atrocity: 1
every: 3
effort: 2
demanded: 1
heatedly: 1
auror: 1
trying: 2
round: 1
happen: 1
most: 2
powerful: 1
eluded: 1
capture: 1
decades: 1
caused: 1
west: 1
temper: 1
rising: 1
pace: 1
infuriating: 1
discover: 1
reason: 1
these: 1
disasters: 1
```

```
able: 1
public: 1
worse: 1
governmentâ: 1
miserably: 1
excuse: 1
barked: 1
positively: 1
stamping: 1
trees: 1
uprooted: 1
roofs: 1
ripped: 1
lampposts: 1
bent: 1
death: 1
eaters: 1
namedâ: 1
suspect: 1
giant: 2
involvement: 2
stopped: 2
tracks: 1
hit: 1
invisible: 1
grimaced: 1
used: 1
giants: 1
wanted: 1
go: 1
grand: 1
effect: 1
misinformation: 1
working: 1
clock: 1
teams: 1
obliviators: 1
modify: 1
memories: 1
running: 1
somerset: 1
disaster: 1
deny: 1
morale: 1
pretty: 1
low: 1
losing: 2
amelia: 3
bones: 3
law: 2
enforcement: 1
her: 1
person: 1
because: 1
she: 4
gifted: 1
witch: 1
real: 1
fight: 1
cleared: 1
throat: 1
seemed: 1
newspapers: 2
momentarily: 1
```

```
diverted: 1
anger: 1
aged: 1
woman: 1
publicity: 1
police: 1
baffled: 1
sighed: 1
locked: 1
gets: 1
us: 1
further: 1
toward: 1
emmeline: 1
vance: 1
hear: 1
here: 1
papers: 1
field: 1
breakdown: 1
order: 1
backyard: 1
barely: 1
listening: 1
swarming: 1
place: 1
attacking: 1
right: 1
center: 1
upon: 1
happier: 1
sentence: 1
unintelligible: 1
wiser: 1
guard: 1
prisoners: 1
cautiously: 1
```

In [24]:

```python
import re
from collections import defaultdict

# Read file with English words which is downloaded from (http://www.gwicks.net/dict
with open(r'/english3.txt', 'r', encoding='ISO-8859-1') as file:
    english_words = set(file.read().lower().split())

#Preprocessing convert to lower cases and removing leadin or trailing punctuation
def preprocess(text):
    text = text.lower()
    words = text.split()
    cleaned_words = [re.sub(r'^[^a-zA-Z\']+', '', re.sub(r'[^a-zA-Z\']+$', '', word
    return ' '.join(cleaned_words)

def mapper(text, english_words):
    words = text.split()
    output = defaultdict(int)
    filtered_words = [word for word in words if "’" not in word]

    # Handling words with hyphens
    for word in filtered_words.copy():
        if '-' in word:
            components = word.split('-')
            if all(component in english_words for component in components):
                filtered_words.remove(word)
```

```
    for word in filtered_words:
        if word not in english_words:
            output[word] += 1

    return output

# Reduce function
def reducer(map_data):
    output = defaultdict(int)
    for word, count in map_data.items():
        output[word] += count
    return output

# Read the file
with open(r'/file2 (1).txt', 'r',encoding='utf-8') as file:
    text_file = file.read()

processed_textFile = preprocess(text_file)

# Running MapReduce
map_data = mapper(processed_textFile, english_words)
reduce_data = reducer(map_data)

#Print Output in Alphabetical order
sorted_dict = dict(sorted(reduce_data.items(), key=lambda x: x[0]))
print(sorted_dict)
```

```
{'auror': 1, 'azkaban': 1, 'delacour': 1, 'diagon': 1, 'down-to\x02earth': 1, 'een
glish': 1, 'eet': 2, 'gringotts': 1, 'im': 1, 'j.k': 10, 'lestrange': 1, 'malfoy':
1, 'rowling': 10, 'seester': 1, 'shh': 1, 'slytherin': 1, 'tchah': 1, 'third\x02el
dest': 1, 'triwizard': 1, 'umbridge': 3, 'voldemort': 3, 'weasley': 12, 'ze': 1,
'zere': 1}
```