

Word count:

```
# Import necessary libraries
from mrjob.job import MRJob
import re
```

```
# Regular expression for finding words in a string
WORD_PATTERN = re.compile(r"\w+")
```

```
# Define a class for counting words
class UniqueWordCount(MRJob):
```

```
    # define the mapper function that reads each line.
```

```
    def mapper(self, _, line):
        for word in WORD_PATTERN.findall(line):
            # Yield each word
            yield word.lower(), 1
```

```
    # define reducer function that receives a word and a list of counts for that word.
```

```
    def reducer(self, word, values):
        # Sum the counts for each word and emit the result
        total_count = sum(values)
        yield word, total_count
```

```
# Run class
```

```
if __name__ == '__main__':
    UniqueWordCount.run()
```

Output:

```
"worries"      1
"would"        15
"wrath"         1
"wrong"         1
"www"18
"year" 3
"years"2
"yes" 1
"yet" 4
"you" 39
"your" 4
"zooming" 1
"ztcprep" 18
"old" 5
"on" 27
"once" 1
```

"one"	9
"only"	5
"open"	3
"opened"	3
"opinion"	1
"or"	4
"orange"	2
"other"	2
"our"	3
"out"	9
"outside"	3
"over"	8
"owl"	1
"p"	11
"pain"	3
"painful"	1
"painfully"	1
"painless"	1
"panic"	1
"paralyzed"	1
"parchment"	1
"parents"	1
"part"	2
"passageway"	1
"passing"	1
"path"	1
"patterned"	1
"peered"	1
"people"	1
"perched"	1
"perfectly"	1
"perform"	2
"performed"	1
"pet"	1
"peter"	1
"petunia"	2
"picked"	1
"picture"	2
"pictures"	1
"pierced"	1
"place"	1
"placed"	1
"plainly"	1
"plan"	3

"planning"	2
"plotting"	2
"pointed"	1
"police"	2
"possible"	1
"potter"	16
"power"	2
"powerful"	2
"powers"	1
"present"	1
"pressed"	2
"privet"	2
"problem"	1
"proceed"	1
"promise"	1
"pronounced"	1
"proved"	1
"put"	2
"putting"	1
"puzzled"	1
"questioned"	1
"questioning"	1
"questions"	1
"quidditch"	1
"quiet"	1
"quietly"	3
"quite"	1
"r"	1
"raised"	1
"ran"	1
"rather"	1
"re"	2
"reaching"	1
"reading"	1
"real"	1
"realized"	1
"really"	1
"rebounded"	1
"recall"	1
"red"	1
"reduced"	1
"reflection"	1
"regrown"	1
"rejoined"	1

"relatives"	1
"remained"	1
"remember"	3
"remorse"	1
"requirement"	1
"respectable"	1
"restlessly"	1
"revealing"	1
"reward"	1
"right"	4
"rise"	1
"robes"	2
"rolls"	1
"room"	12
"rot"	1
"rotting"	1
"roughly"	1
"round"	1
"rowing"	11
"rug"	4
"run"	1
"running"	2
"rushing"	1
"s"	12
"said"	16
"same"	1
"sat"	3
"saturday"	1
"saw"	1
"say"	1
"scar"	10
"wasn"	4
"watched"	2
"water"	2
"watery"	1
"wavered"	1
"way"	1
"wayside"	1
"we"	3
"welcome"	1
"well"	3
"went"	4
"were"	16
"what"	10

"when"	4
"where"	5
"wherever"	1
"which"	6
"whimper"	1
"whispered"	1
"whispers"	1
"white"	1
"who"	5
"whoever"	1
"whose"	1
"why"	3
"wide"	2
"wife"	2
"will"	10
"window"	2
"wire"	1
"wish"	2
"witchcraft"	1
"witches"	3
"with"	19
"without"	4
"wizard"	7
"wizarding"	2
"wizardry"	1
"wizards"	4
"woke"	1
"woman"	1
"wooden"	1
"word"	1
"words"	1
"worked"	1
"world"	3
"wormtail"	24
"fright"	1
"from"	8
"frowning"	1
"frozen"	1
"fulfill"	1
"full"	2
"fuss"	1
"g"	11
"gaining"	1
"gap"	1

"gave"	1
"gigantic"	1
"give"	2
"glasses"	1
"go"	2
"goal"	1
"goblet"	11
"going"	3
"gone"	4
"good"	1
"graying"	1
"green"	2
"grip"	1
"gripping"	1
"ground"	1
"growing"	1
"grown"	1
"grunting"	1
"had"	62
"hadn"	2
"hair"	2
"half"	1
"hand"	4
"hands"	5
"happened"	1
"hard"	2
"harry"	51
"began"	1
"behind"	1
"being"	2
"below"	1
"beneath"	1
"bertha"	4
"beside"	2
"best"	2
"beyond"	1
"birthday"	1
"bizarre"	1
"black"	2
"blame"	1
"blocking"	1
"body"	1
"bolt"	2
"bones"	1

"book"	4
"bothered"	1
"bothering"	1
"bottle"	1
"box"	1
"boy"	4
"boys"	1
"brains"	1
"braver"	1
"breath"	1
"breathing"	1
"bright"	2
"brilliance"	1
"broken"	1
"broomstick"	2
"broomsticks"	1
"brought"	1
"brutus"	1
"bryce"	1
"burning"	1
"but"	16
"by"	11
"cage"	1
"called"	2
"calling"	2
"came"	2
"can"	1
"cannons"	1
"case"	1
"cast"	1
"castle"	1
"cat"	1
"caught"	1
"cauldron"	1
"center"	1
"century"	1
"chair"	6
"chance"	1
"changed"	1
"charms"	1
"clatter"	1
"clear"	1
"clearer"	1
"cloak"	1

"close"	1	
"closed"	2	
"closely"	2	
"closer"	2	
"cold"	12	
"com"	18	
"come"	6	
"coming"	3	
"community"	1	
"concentrated"		1
"condition"	1	
"confide"	1	
"confused"	1	
"continuing"	1	
"coolly"	1	
"corridor"	2	
"could"	7	
"couldn't"	4	
"counting"	1	
"courage"	1	
"cousin"	1	
"crackling"	1	
"creak"	1	
"creature"	1	
"creep"	1	
"criminal"	1	
"crossed"	2	
"cruel"	1	
"crumpled"	1	
"cube"	1	
"cupped"	1	
"curled"	1	
"curse"	2	
"curtains"	3	
"curving"	1	
"cut"	2	
"danger"	1	
"dangerous"	1	
"dark"	3	
"darkened"	1	
"darkness"	2	
"days"	1	
"dead"	2	
"death"	1	



"decided"	1	
"decision"	1	
"deeply"	1	
"defiantly"	1	
"definitely"	2	
"deny"	1	
"desk"	1	
"despised"	1	
"details"	1	
"diamond"	1	
"didn"	4	
"died"	1	
"dim"	1	
"dimly"	1	
"disappearance"		1
"disbanded"	1	
"disconcerting"		1
"discover"	1	
"disposed"	1	
"distract"	1	
"do"	8	
"dog"	1	
"don"	5	
"done"	2	
"door"	3	
"down"	2	
"drawing"	1	
"dream"	1	
"dreaming"	1	
"dreams"	1	
"drew"	2	
"drive"	2	
"dry"	2	
"dudley"	3	
"dursleys"	2	
"dust"	1	
"e"	11	
"early"	1	
"edge"	1	
"eleven"	1	
"eleventh"	1	
"else"	2	
"empty"	1	
"end"	2	

"english"	1	
"enough"	2	
"entirely"	1	
"escape"	1	
"essential"	1	
"even"	5	
"ever"	3	
"every"	1	
"everyone"	2	
"everything"	1	
"exactly"	1	
"examined"	1	
"except"	1	
"expected"	1	
"expecting"	2	
"explained"	1	
"extent"	1	
"extinguished"	1	
"extracted"	1	
"extraordinary"	1	1
"eyes"	4	
"face"	7	
"facing"	1	
"faint"	1	
"faithful"	3	
"fall"	1	
"fallen"	1	
"famous"	1	
"fang"	1	
"far"	1	
"fast"	1	
"father"	1	
"fear"	1	
"feel"	1	
"feet"	2	
"fell"	2	
"felt"	3	
"fifty"	2	
"filtering"	1	
"find"	3	
"finger"	1	
"fingers"	2	
"fire"	13	
"firelight"	1	

"firmer"	1
"first" 1	
"fit" 2	
"flames"	1
"flash" 1	
"flat" 1	
"fled" 1	
"floor" 5	
"flung" 1	
"flying"	1
"focus" 1	
"follow"	1
"followed"	1
"followers"	2
"following"	1
"foot" 3	
"footsteps"	1
"for" 20	
"forehead"	4
"form" 1	
"formed"	1
"forward"	1
"found"	3
"fourteen"	1
"fourth"	1
"frank" 23	
"stomach"	1
"stood"	3
"straight"	1
"strange"	1
"stranger"	1
"street"	3
"stroke"	1
"stupid"	1
"suburban"	1
"such" 1	
"sudden"	1
"suddenly"	2
"sulky" 1	
"sullenness"	1
"summer"	1
"supposed"	1
"surely"	1
"surprise"	1

"survey"	1
"survived"	1
"sweat"	2
"swish"	1
"switched"	1
"swung"	1
"t" 19	
"table" 2	
"tail" 1	
"taken"	1
"talk" 1	
"talking"	2
"task" 1	
"telephone"	1
"tell" 2	
"telling"	2
"terrified"	1
"terror"	2
"than" 5	
"that" 38	
"the" 181	
"their" 3	
"them"7	
"then" 7	
"there"14	
"they" 10	
"thick" 1	
"thing" 2	
"things"	1
"think" 2	
"this" 7	
"though"	10
"thought"	4
"three"1	
"threshold"	1
"thrill" 1	
"through"	6
"throwing"	1
"tightly"	1
"time" 4	
"tip" 1	
"to" 71	
"told" 2	
"tonight"	1

"too"	2	
"took"	2	
"toward"	1	
"trace"	1	
"track"	1	
"transfixed"	1	
"travesty"	1	
"trembling"	1	
"tremendous"	1	
"triangular"	1	
"trickling"	1	
"tried"	3	
"trouble"	1	
"true"	1	
"trunk"	1	
"truth"	1	
"trying"	4	
"turn"	3	
"turned"	4	
"twelve"	1	
"two"	3	
"ugly"	1	
"unavoidable"	1	
"uncle"	2	
"under"	1	
"underage"	1	
"understand"	1	
"undulating"	1	
"unfortunately"	1	1
"unless"	1	
"unnoticed"	1	
"untidy"	1	
"until"	1	
"untroubled"	1	
"unusual"	2	
"up"	6	
"upon"	3	
"us"	1	
"use"	2	
"used"	2	
"useful"	3	
"useless"	1	
"usually"	1	
"vanished"	1	

"ve"	2
"venomous"	1
"vernon"	2
"very"	5
"village"	1
"visited"	1
"vivid"	1
"voice"	22
"voldemort"	19
"walked"	2
"walking"	4
"walls"	1
"wand"	2
"want"	2
"wanted"	1
"war"	1
"wardrobe"	1
"was"	64
"school"	1
"score"	1
"scrambled"	1
"scream"	1
"screaming"	1
"screwed"	1
"second"	5
"seconds"	1
"secret"	1
"secure"	1
"see"	4
"seem"	1
"seemed"	2
"seen"	1
"seizure"	1
"servant"	5
"shadows"	1
"shaken"	1
"shaking"	1
"shaped"	2
"she"	5
"shock"	1
"shook"	1
"short"	1
"shut"	1
"sight"	2

"silence"	1
"silky"	1
"sitting"	2
"skin"	1
"skinny"	1
"slightly"	2
"slipped"	1
"slippery"	1
"slithering"	1
"sliver"	1
"slowly"	1
"small"	4
"smaller"	1
"snagged"	1
"snake"	7
"snakes"	1
"snapped"	1
"snore"	1
"snowy"	1
"so"	5
"some"	3
"somebody"	2
"someone"	3
"something"	4
"sort"	2
"sound"	1
"sounded"	2
"sounding"	1
"source"	1
"spasm"	1
"speak"	1
"speaker"	1
"speaking"	1
"spectacular"	1
"speech"	1
"spellbooks"	1
"spidery"	1
"spitting"	2
"spoil"	1
"spoke"	1
"sport"	1
"spot"	1
"st"	1
"stair"	1

"standing"	1
"stared"	3
"start" 1	
"started"	1
"starting"	1
"stayed"	1
"steadier"	1
"steadily"	1
"steady"	1
"stick" 4	
"still" 7	
"stinging"	1
"its" 4	
"j" 11	
"jorkins"	3
"journey"	1
"jumped"	1
"just" 2	
"k" 11	
"keep" 1	
"kill" 4	
"killed"4	
"killing"	1
"kind" 1	
"knack"	1
"knew"5	
"know"	3
"knows"	4
"lamp" 2	
"large" 2	
"last" 4	
"laugh"	2
"laughable"	1
"lay" 3	
"least" 1	
"legs" 2	
"let" 1	
"level" 1	
"liar" 1	
"lie" 1	
"life" 2	
"lifted" 2	
"light" 3	
"lightning" 4	



"like"	5
"liked"	1
"limped"	1
"listened"	1
"listening"	2
"lit"	2
"littered"	1
"live"	1
"lived"	1
"living"	2
"ll"	1
"long"	7
"look"	2
"looked"	4
"lord"	10
"lost"	1
"lot"	1
"loudly"	1
"loyalty"	1
"lurking"	1
"m"	1
"made"	3
"madman"	1
"magic"	3
"making"	1
"man"	14
"manners"	2
"many"	2
"another"	1
"any"	4
"anyone"	1
"anything"	4
"approach"	1
"apt"	1
"are"	5
"argument"	1
"arm"	2
"armchair"	2
"around"	4
"arrived"	2
"as"	29
"asked"	1
"asking"	1
"asleep"	3

"assorted"	1
"at" 16	
"attended"	1
"attracting"	1
"audible"	1
"aunt" 2	
"awake"	1
"aware"	2
"away" 3	
"awkward"	1
"awoke"	1
"awoken"	3
"back" 10	
"balding"	1
"ball" 2	
"barely"	2
"be" 14	
"became"	1
"because"	3
"beckoned"	1
"become"	1
"becoming"	1
"bed" 5	
"bedroom"	3
"bedside"	2
"been" 23	
"before"	7
"has" 3	
"hated"	1
"have" 14	
"having"	1
"he" 61	
"head" 4	
"heads"	1
"hear" 3	
"heard"	7
"hearth"	3
"hedwig"	1
"help" 1	
"her" 7	
"here" 3	
"hidden"	1
"hide" 1	
"high" 2	

"him"	20
"himself"	4
"his"	68
"hiss"	1
"hissed"	1
"hissing"	2
"hit"	1
"hoarse"	1
"hogwarts"	6
"hold"	2
"holidays"	1
"honor"	1
"hoop"	1
"horrible"	1
"horrified"	1
"horror"	1
"hot"	2
"hours"	1
"house"	5
"how"	1
"however"	2
"hundred"	1
"hurt"	1
"hurting"	1
"i"	38
"ice"	1
"idea"	4
"if"	13
"impossible"	3
"in"	41
"incredibly"	2
"incurably"	1
"indeed"	2
"information"	2
"injuries"	1
"injury"	1
"inns"	1
"inside"	5
"inspiration"	1
"instead"	1
"insult"	1
"interest"	1
"interesting"	1
"into"	7

"invaluable"	1
"invite"	1
"is"	8
"it"	43
"11"	1
"12"	1
"13"	1
"14"	1
"15"	1
"16"	1
"17"	1
"18"	1
"19"	1
"20"	1
"21"	1
"a"	74
"able"	1
"about"	9
"abruptly"	1
"absences"	1
"absurd"	1
"accidents"	1
"according"	1
"action"	1
"added"	1
"after"	1
"afterward"	1
"again"	7
"ah"	1
"airborne"	1
"alarm"	1
"alive"	1
"all"	8
"allow"	1
"allowed"	1
"almost"	1
"alone"	1
"along"	1
"already"	1
"always"	2
"am"	5
"amused"	1
"amusement"	2
"an"	10

"ancient"	1
"and"	86
"master"	2
"me"	6
"mean"	1
"means"	2
"meant"	1
"memory"	3
"men"	2
"mentally"	1
"merest"	1
"met"	1
"might"	2
"miles"	1
"mine"	1
"ministry"	3
"miraculously"	1
"mirror"	1
"mirthless"	1
"misty"	1
"mixture"	1
"modified"	1
"moment"	2
"more"	10
"morning"	1
"most"	1
"mother"	1
"mouth"	2
"move"	1
"movement"	1
"moving"	1
"much"	3
"muggle"	5
"muggles"	1
"murder"	4
"murders"	1
"must"	3
"muttered"	1
"my"	14
"myself"	1
"nagini"	3
"name"	2
"nearer"	1
"need"	2

```
"neither"    1
"never"      6
"news"       2
"next"       1
"nicknamed"  1
"night"      3
"no"         5
"nobody"     2
"noises"     2
"normal"     1
"nose"       1
"not"        18
"nothing"    2
"now"        10
"number"     1
"of"         64
```

Non-eng :

```
from mrjob.job import MRJob
import enchant
```

```
# create a dictionary for US English words
d = enchant.Dict("en_US")
```

```
class NonEnglishWordsCount(MRJob):
```

```
    # the mapper function reads each line and splits it into word
```

```
    def mapper(self, _, line):
        for word in line.split():
            if not d.check(word):
                yield word.lower(), 1
```

```
    # the reducer function receives a word and a list of counts for that word
```

```
    def reducer(self, word, counts):
        yield word, sum(counts)
```

```
if __name__ == '__main__':
    NonEnglishWordsCount.run()
```

Output:

```
gryffindor,"
"\u2013"
1 11
```

"\u2014"  
10  
"\u2014\u201d" 2  
"\u2018fortuna" 1  
"\u2018sibyll" 1  
"\u201c"  
1  
"\u201ca"  
1  
"\u201caha!\u201d"  
1  
"\u201call"  
2  
"\u201cand"  
1  
"\u201canyway," 1  
"\u201care"  
1  
"another," "applause," 1  
"appointment, \u201d"  
1  
"arithmancy,' 1  
"around,"  
1  
"azkaban"  
12  
"azkaban,  
1  
"back," 2  
"be,\u201d"  
1  
"beard,"  
1  
"bed," 1  
"beds," 1  
"been," 1  
"begin! \u201d" 1  
"blandly,"  
1  
"book?\u201d" 1  
"braggart!" 1  
"business.\u201d"  
1  
"cadogan!\u201d"

1  
"castle,"  
1  
"ceiling,"  
1  
"charge!\u201d" 1  
"cheerful,"  
1  
"clapping," 1  
"cloaks, \u201d" 1  
"harry, \u201d" 1  
"harry?\u201d" 1  
"hastily,"  
1  
"hat, 1  
"he," 1  
"heart,"  
1  
"hermione"  
13  
"hermione," "hermione."  
"him," 1  
4  
H  
2  
"him, \u201d"  
1  
"him;" 1  
"himself, \u201d"  
1  
"hogwarts!" 1  
"hones\u2019y."  
1  
"however,"  
2  
"hufflepuff," 1  
"importance, \u201d"  
1  
"insides,"  
1  
"it,' 1  
"j.k." 11  
"job," 1  
"kettleburn" 1



"kettleburn," 1

"knaves,"

1