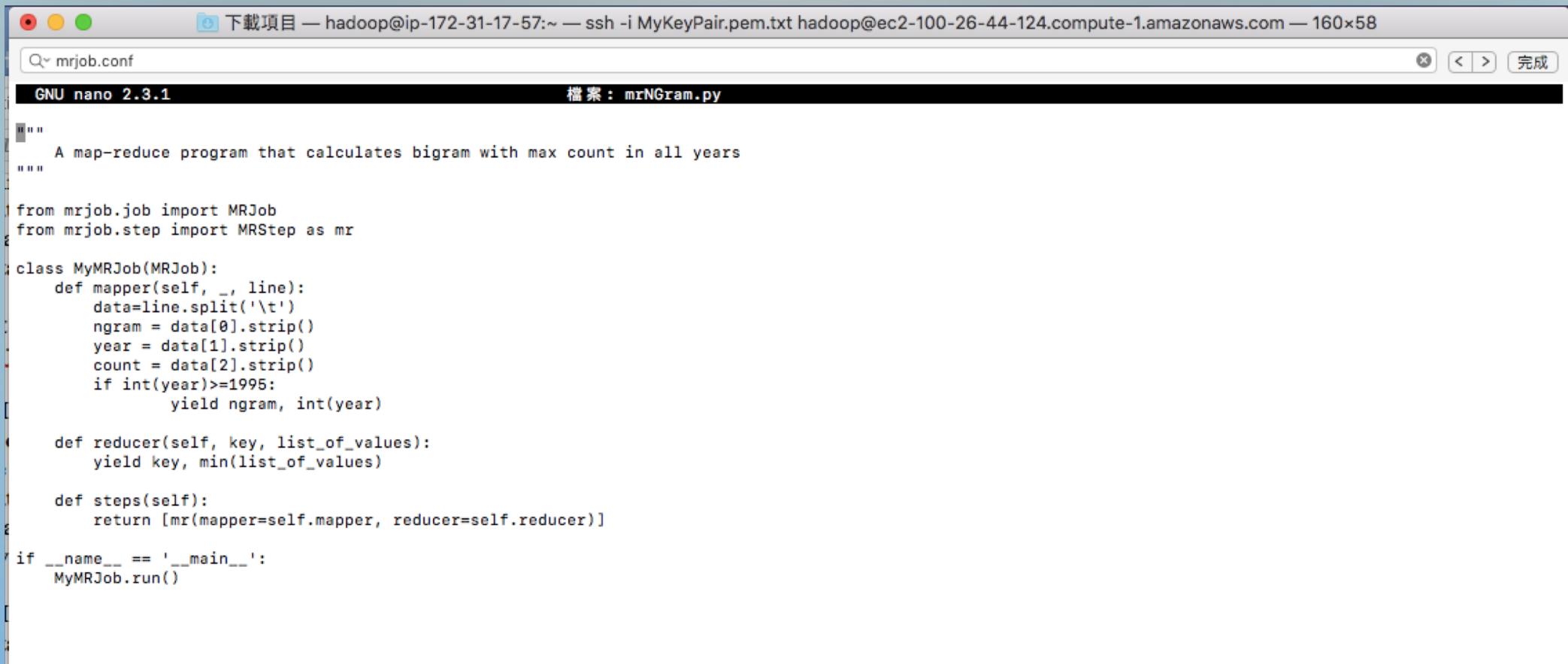


EXERCISE-1



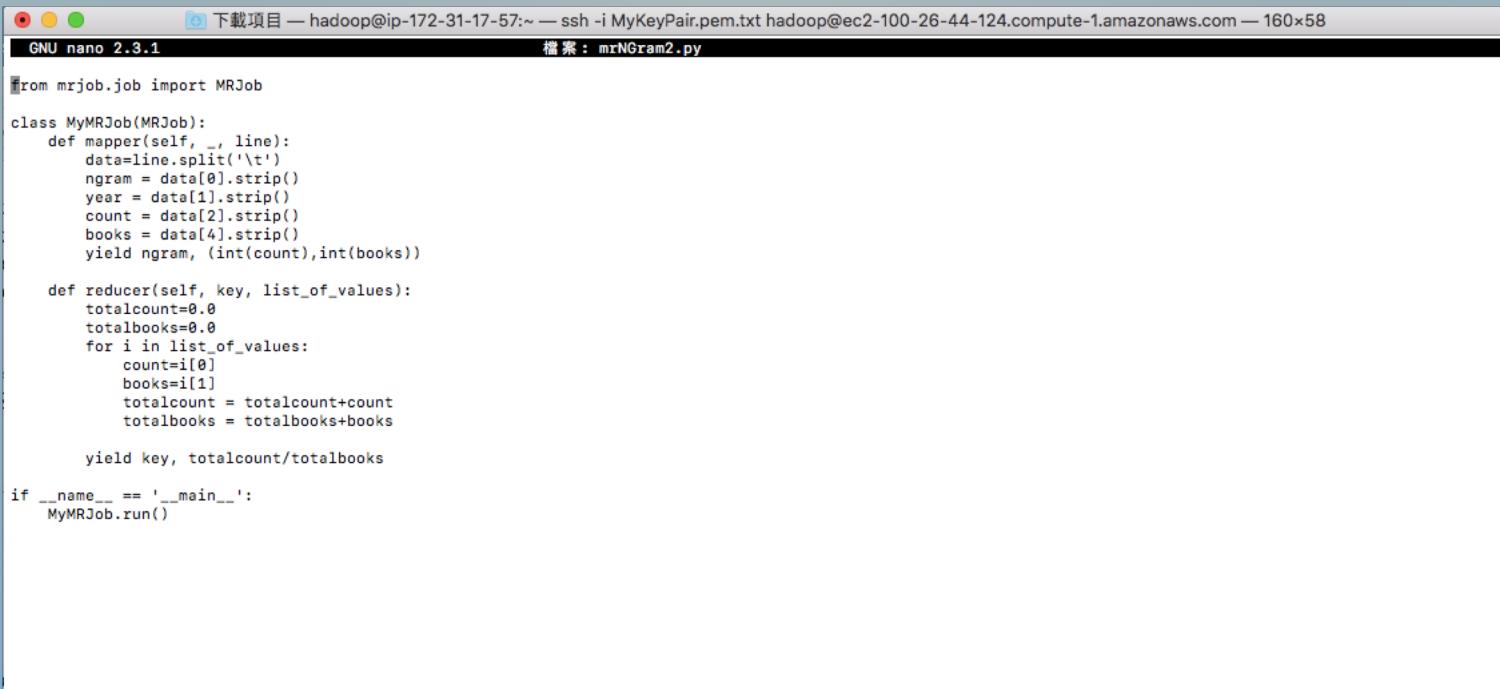
The screenshot shows a terminal window titled "下載項目 — hadoop@ip-172-31-17-57:~ — ssh -i MyKeyPair.pem.txt hadoop@ec2-100-26-44-124.compute-1.amazonaws.com — 160x58". The window displays a file named "mrjob.conf" being edited with "GNU nano 2.3.1". The code is a Python script for a map-reduce job:

```
"""  
A map-reduce program that calculates bigram with max count in all years  
"""  
  
from mrjob.job import MRJob  
from mrjob.step import MRStep as mr  
  
class MyMRJob(MRJob):  
    def mapper(self, _, line):  
        data=line.split('\t')  
        ngram = data[0].strip()  
        year = data[1].strip()  
        count = data[2].strip()  
        if int(year)>=1995:  
            yield ngram, int(year)  
  
    def reducer(self, key, list_of_values):  
        yield key, min(list_of_values)  
  
    def steps(self):  
        return [mr(mapper=self.mapper, reducer=self.reducer)]  
  
if __name__ == '__main__':  
    MyMRJob.run()
```

Output:

"! \$17.95"	1995	"! 13.5"	1995	"& Joule"	1996	"\$1200000 ."	1995
"! 1791"	1995	"! 6ng"	2004	"& Karlson"	1995	"\$125 per"	1995
"! 1847"	1996	"! Advocacy"	1996	"& Katz"	1995	"\$12650 ."	1995
"! Acknowledge"	1995	"! Albers"	1997	"& Keg"	1995	"\$129 billion"	1995
"! Analytical"	1995	"! Anorexia"	1995	"& Longmire"	1996	"\$135000 on"	1995
"! Anna"	1995	"! Antipas"	1995	"& Meister"	1995	"\$13950 ."	1995
"! Autograph"	1995	"! Assholes"	1995	"& Menus"	1995	"\$14.95 How"	1995
"! Bobo"	1995	"! BEATRICE"	1995	"& Mindy"	1995	"\$14500 ;"	1995
"! Budapest"	1995	"! BEAUTY"	1995	"& Modernisation"	1995	"\$14763 for"	1995
"! Chagrined"	1997	"! BTW"	1995	"& Nich"	1996	"\$14800 ("	1995
"! Communion"	1995	"! Blossoming"	1996	"& Ownership"	1995	"\$150 \$170"	1997
"! Cossack"	1996	"! Books"	1995	"& PROVINCIAL"	1996	"\$150 weekly"	1995
"! Drowned"	1995	"! Bowers"	1995	"& Parmentier"	1995	"\$1500.00 The"	1999
"! Edict"	1996	"! Byzantium"	1996	"& Percentages"	1996	"\$15000 \$5000"	1995
"! Effective"	1995	"! CBS"	1995	"& Perez"	1995	"\$153.2 million"	1995
"! Explanlation"	1995	"! Cara"	1996	"& Plaskitt"	1995	"\$175 plus"	1996
"! Girty"	1995	"! Commodity"	1995	"& Pr"	1995	"\$19300)"	1996
"! Gladden"	1996	"! Cowardly"	1996	"& Racusin"	1995	"\$194000)"	1998
"! Gleam"	1995	"! Daedalus"	1995	"& Raviv"	1995	"\$2 does"	2000
"! Handcuffs"	1995	"! ESSAY"	1996	"& STEIN"	1995	"\$2 out"	1995
"! Hora"	1996	"! Essentials"	1995	"& Sax"	1995	"\$2.0 M"	1995
"! Horseback"	1996	"! Famous"	1995	"& Sculptors"	1995	"\$2.00 Electrotyping"	2007
"! Ineffuctual"	2000	"! Fate"	1995	"& Skalak"	1995	"\$2.45 for"	1995
"! Kathe"	1995	"! Fidon"	2002	"& Snowboarding"	1998	"\$2.50 MORRIS"	2007
"! Lamb's"	1998	"! Flip"	1995	"& Standen"	1995	"\$20000 Interest"	1995
"! MISCELLANEOUS"	1996	"! Flourishing"	1996	"& Stepper"	1995	"\$200000 before"	1995
"! Medico"	1995	"! GUARD"	1995	"& Szalacha"	2001	"\$21 *"	1995
"! NED"	1995	"! Ginsberg"	1995	"& THEODORE"	1997	"\$24.95 /"	1995
"! Nico"	1996	"! Idle"	1995	"& Teubner"	1995	"\$240000 annually"	1996
"! Peleg"	1996	"! Idols"	1996	"& Tornatzky"	1995	"\$2400000 were"	1999
"! Postage"	1995	"! Improvisation"	1995	"& Toulouse"	1995	"\$24000000 in"	1996
"! Preferring"	1995	"! Injure"	1995	"& Trattner"	1995	"\$245 ;"	1995
"! REMARKS"	1996	"! Johanna"	1995	"& Tremoulet"	1996	"\$25.00 when"	1995
"! Recipe"	1995	"! Lancashire"	1995	"& Usages"	2003	"\$25000 b"	1997
"! Risking"	1995	"! Liberalism"	1995	"& Wandel"	1996	"\$25000 homestead"	1996
"! Ronald"	1995	"! Lifted"	1996	"& Welling"	1995	"\$25000 profit"	1995
"! September"	1995	"! Mayhap"	1995	"& YOU"	1995	"\$250000 :"	1995
"! Slaughter"	1995	"! McLeod"	1998	"& Zarrow"	1995	"\$26400 ;"	1996
"! Socrates"	1995	"! Morelli"	1996	"& Zeichner"	1995	"\$27800000 ."	1995
"! Stores"	1995	"! Neapolitan"	2000	"& aging"	1995	"\$27800000 in"	2001
"! Tanto"	1995	"! PHILADELPHIA"	1995	"& aviation"	1997	"\$28.50 ."	1995
"! Tertullian"	1995	"! PRAXINOA"	1997	"& buy"	1995	"\$29000 or"	1995
"! Thrice"	1995	"! Partner"	1995	"& certain"	1995	"\$3 profit"	1995
"! Torpedoes"	1995	"! Probably"	1995	"& coasts"	1995	"\$3.00 21"	2000
"! Trifle"	1996	"! Programmes"	1998	"& convince"	1995	"\$3.50 Volume"	2007
"! Uberall"	1995	"! Radius"	1995	"& density"	1995	"\$3.75 plus"	1996
"! Unutterably"	2001	"! SHEEP"	1995	"& depart"	1995	"\$300 .00"	2007
"! Whoy"	1996	"! Sauerkraut"	1995	"& detached"	1995	"\$3000 sale"	1996
"! abscess"	1995	"! Spawn"	1995	"& drastic"	1995	"\$300000 advance"	1995
"! blamed"	1995	"! Spin"	1995	"& drill"	1995	"\$312 million"	1995
"! celebrations"	1996	"! Striker"	1996	"& embryo"	1995	"\$335 /"	1995
"! compound"	1996	"! Suite"	1995	"& escapes"	1998	"\$34000000 in"	1997
"! destroyer"	1997	"! THACKERAY"	2003	"& forth"	1995	"\$351 for"	1997
"! exceed"	1996	"! Talia"	1996	"& hardworking"	1995	"\$4 -"	1995
"! exemplified"	2001	"! Tonnage"	2000	"& inert"	1995	"\$4 ;"	1995
"! extended"	1996	"! Vere"	1995	"& initial"	1995	"\$4.12 million"	1996
"! fickle"	1995	"! Wait"	1995	"& intensely"	1995	"\$40 \$100"	1995
"! filtering"	1995	"! alio"	1995	"& lending"	1995	"\$400 \$600"	1995
"! implementation"	1995	"! amit"	2006	"& load"	1995	"\$42.22 per"	1995
"! lier"	1996	"! appeared"	1995	"& matrimonial"	1996	"\$4300 ("	1995
"! melt"	1997	"! believing"	1997	"& merchandise"	1995	"\$4300000 ;"	2002
"! metres"	1995	"! centered"	1996	"& perverting"	1995	"\$435.5 million"	1996
"! nate"	1995			"& nrit"	2008	"\$465000 for"	1995

EXERCISE-2



The screenshot shows a terminal window titled "下載項目 — hadoop@ip-172-31-17-57:~ — ssh -i MyKeyPair.pem.txt hadoop@ec2-100-26-44-124.compute-1.amazonaws.com — 160x58". The window title bar also displays "GNU nano 2.3.1" and "檔案: mrNGram2.py". The terminal content is a Python script named "mrNGram2.py" which defines a class "MyMRJob" that extends "MRJob". The "mapper" method splits each line by '\t', extracts the first four fields (ngram, year, count, books), and yields them as a tuple. The "reducer" method initializes totalcount and totalbooks to 0.0, then iterates over the list of values for each key, summing up the counts and books, and finally yielding the key along with the average count per book. The script concludes with an "if __name__ == '__main__': MyMRJob.run()" block.

```
from mrjob.job import MRJob

class MyMRJob(MRJob):
    def mapper(self, _, line):
        data=line.split('\t')
        ngram = data[0].strip()
        year = data[1].strip()
        count = data[2].strip()
        books = data[4].strip()
        yield ngram, (int(count),int(books))

    def reducer(self, key, list_of_values):
        totalcount=0.0
        totalbooks=0.0
        for i in list_of_values:
            count=i[0]
            books=i[1]
            totalcount = totalcount+count
            totalbooks = totalbooks+books

        yield key, totalcount/totalbooks

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

Output:

!"\$17.95"	1.0909090909090908	!"1771"	1.048780487804878	!"09"	1.1996830427892236	!"Laurie"	1.373134328358209
!"1791"	1.457142857142857	!"431"	1.0309278350515463	!"960"s"	1.10666666666666667	!"MITCHELL"	1.4578313253012047
!"1847"	1.2397260273972603	!"5.7"	1.03108800829015545	!"AS"	1.2478238607270866	!"McDowell"	1.0347826086956522
!"Acknowledge"	1.0487062404870624	!"ANNE"	3.0885860306643953	!"Barron"	1.7906976744186047	!"Medication"	1.0085470085470085
!"Analytical"	1.1967213114754098	!"Abort"	1.4343434343434343	!"Bellarmine"	1.2333333333333334	!"Melissa's"	1.14
!"Anna"	1.6472763125462164	!"Abuela"	1.5238095238095237	!"Berthold"	1.4210526315789473	!"Memorials"	1.0133333333333334
!"Autograph"	1.0555555555555556	!"Alphonso"	1.310077519379845	!"Bienvenidos"	1.0408163265306123	!"Mod"	1.2736842105263158
!"Bobo"	1.3838383838383839	!"Annabelle"	1.685483870967742	!"Blowed"	1.0129870129870129	!"Momentarily"	1.073482428115016
!"Budapest"	1.08843537414966	!"Bile"	1.00847457262711864	!"Cardiovascular"	1.1047619047619048	!"Mustafa"	1.2682926829268293
!"Chagrined"	1.0273972602739727	!"Blocking"	1.007936507936508	!"Chambers"	1.0754716981132075	!"Named"	1.0353982300884956
!"Communion"	1.0555555555555556	!"Branch"	1.0555555555555556	!"Che"	1.4455300439667806	!"Napoleonic"	1.0701754385964912
!"Cossack"	1.0632911392405062	!"Chic"	1.12987012987013	!"Chopper"	1.0434782608695652	!"Personal"	1.124856815578465
!"Drowned"	1.078512396694215	!"Clock"	1.1262626262626263	!"Claudius"	1.3905109489051095	!"Pickpockets"	1.0
!"Edict"	1.075	!"Corrode"	1.077922077922078	!"Conjecture"	1.0258620689655173	!"Psal"	1.5592105263157894
!"Effective"	1.046405823475887	!"Cromwell's"	1.1348314606741574	!"Custom"	1.0675496688741721	!"ROSlE"	5.883116883116883
!"Explanation"	1.4513064133016627	!"Cyclones"	1.0	!"David's"	1.125	!"Rejoicing"	1.0321428571428573
!"Girty"	1.0746268656716418	!"Days"	1.0544959128065394	!"Disconcerted"	1.0222222222222221	!"Relieve"	1.0468431771894093
!"Gladden"	1.014018691588785	!"Distinct"	1.0178571428571428	!"Emphatically"	1.0091324200913243	!"Rikk"	1.0
!"Gleam"	1.0588235294117647	!"En"	1.2743137919349372	!"Gottlieb"	1.641711229946524	!"Rippling"	1.0384615384615385
!"Handcuffs"	1.0246913580246915	!"Epitaph"	1.0476190476190477	!"Hare"	1.3755795981452859	!"Send"	1.2078906746572233
!"Hora"	1.2916666666666667	!"Flashes"	1.028436018957346	!"Ire"	1.5714285714285714	!"Shoulda"	1.036144578313253
!"Horseback"	1.0344827586206897	!"Fro"	1.028735632183908	!"Lecture"	1.0585774058577406	!"Sluggish"	1.0
!"Ineffuctual"	1.0	!"Giggling"	1.0	!"MOLLY"	3.3253968253968256	!"Snuff"	1.0820095522388059
!"Kathe"	1.325	!"Glories"	1.075	!"Markheim"	1.0135135135135136	!"Soul"	1.11651728553137
!"Lamb's"	1.0074074074074073	!"Handled"	1.09375	!"NEBRASKA"	1.1904761904761905	!"Storyteller"	2.370967741935484
!"MISCELLANEOUS"	1.0810810810810811	!"Heartbreak"	1.1884057971014492	!"Oa"	1.043103448275862	!"Tariffs"	1.0222222222222221
!"Medico"	1.0851063829787233	!"Hero"	1.300411522633745	!"Offspring"	1.0192307692307692	!"Tractors"	1.1372549019607843
!"NED"	3.314814814814815	!"Huayna"	1.0212765957446808	!"Originator"	1.125	!"Transmute"	1.025
!"Nico"	1.7162162162162162	!"ICHABOD"	2.15	!"Pathos"	1.06666666666666667	!"Ui"	1.52803738317757
!"Peleg"	1.146067415730337	!"INFORMATION"	1.1452282157676348	!"Photograph"	1.172077922077922	!"Undergraduates"	1.3095238095238095
!"Postage"	1.0416666666666667	!"Lawson"	1.1415094339622642	!"Praising"	1.0754716981132075	!"Une"	1.1580161476355249
!"Preferring"	1.0091743119266054	!"Lazily"	1.0	!"Ree"	1.1587301587301588	!"Unending"	1.0632911392405062
!"REMARKS"	1.4880952380952381	!"Lenses"	1.0227272727272727	!"Rerum"	1.0204081632653061	!"Wal"	1.5772200772200773
!"Recipe"	1.5968586387434556	!"MELISSA"	2.2266666666666666	!"STAR"	1.8038277511961722	!"Wedge"	1.0
!"Risking"	1.0256410256410255	!"Maimonides"	1.0638297872340425	!"STAY"	1.178021978021978	!"criterion"	1.0
!"Ronald"	1.1612483745123536	!"Mathew"	1.135593202339831	!"Saturn"	1.1692307692307693	!"demonic"	1.0265486725663717
!"September"	1.32835201328352	!"Myrna"	1.1411764705882352	!"Screening"	1.010989010989011	!"disait"	1.1079136690647482
!"Slaughter"	1.1428571428571428	!"Noses"	1.1320754716981132	!"Sess"	1.11475409836056556	!"jose"	1.0
!"Socrates"	1.492323439092834	!"Obscenity"	1.0135135135135136	!"Shades"	1.0285027472527473	!"kiss"	1.0476190476190477
!"Stores"	1.3511904761904763	!"Pavel"	1.5160142348754448	!"Shakes"	1.1139240506329113	!"leges"	1.0161290322580645
!"Tanto"	1.0236686390532543	!"Perceptions"	1.0120481927710843	!"Sherlock"	1.1256281490735176	!"lei"	1.0
!"Tertullian"	1.2627737226277371	!"Polyester"	1.024390243902439	!"Snarl"	1.392156862745098	!"mariners"	1.064516129032258
!"Thrice"	1.0562080536912752	!"Principalities"	1.0238095238095237	!"Somos"	1.1717171717171717	!"ministraton"	1.0
!"Torpedoes"	1.0679611650485437	!"Proclamation"	1.0551724137931036	!"Strings"	1.0609756097560976	!"mirth"	1.0
!"Trifle"	1.025	!"Prussians"	1.0	!"Tempestuous"	1.2040816326530612	!"movements"	1.0382165605095541
!"Uberall"	1.0217391304347827	!"Realistic"	1.0432098765432098	!"Timmins"	1.0	!"paragraph"	1.0775862068965518
!"Unutterably"	1.0	!"Recent"	1.0317535545023697	!"UH"	1.0833333333333333	!"rt"	1.0448383733055266
!"Whoy"	1.0961538461538463	!"Renegade"	1.0636942675159236	!"Unsigned"	1.069767441860465	!"same"	1.037190082644628
!"abscess"	1.1003236245954693	!"Ricky's"	1.0476190476190477	!"Vindictive"	1.0	!"seal"	1.0696517412935322
!"blamed"	1.0119047619047619	!"Roaring"	1.0434782608695652	!"Virgo"	1.3299492385786802	!"skills"	1.101522842639594
!"celebrations"	1.0	!"SCAN"	2.05	!"Water's"	1.0064102564102564	!"subscription"	1.0344827586206897
!"compound"	1.0060240963855422	!"SHAKESPEARE"	1.3805309734513274	!"Wbat"	1.0135135135135136	!"tcr"	1.037037037037037
!"destroyer"	1.0380952380952382	!"SUNDAY"	1.528	!"XT"	1.0140845070422535	!"union"	1.0349794238683128
!"exceed"	1.3802083333333333	!"Sands"	1.191304347826087	!"Vu043e"	1.1374045801526718	!"worksheet"	2.65
!"exemplified"	1.0178571428571428	!"Sempe"	1.218045112781955	!"agitation"	1.0396039603960396	#"1247"	1.3495145631067962
!"extended"	1.0536912751677852	!"Shrapnel"	1.0384615384615385	!"bill"	1.1973684210526316	#"30"	1.7687055742611297
!"fickle"	1.1094890510948905	!"Solemnly"	1.0541516245487366	!"concerts"	1.0222222222222221	#"65"	1.5675057208237986
!"filtering"	1.024390243902439	!"Sorted"	1.0384615384615385	!"dealing"	1.016260162601626	#"984"	1.1951219512195121
!"implementation"	1.043859649122807	!"Stephanie's"	1.1162790697674418	!"deceased"	1.0135135135135136	#"DC"	1.2549019607843137
!"lier"	1.0555555555555556	!"Stout"	1.054014598540146	!"depreciation"	1.0	#"change"	1.206896551724138
!"melt"	1.0	!"TAL"	1.0	!"diversi"	1.0704225352112675	#"est"	1.3
!"metres"	1.0941176470588236	!"Tarleton"	1.06	!"erythema"	1.3043478260869565	#"0.03"	1.0952380952380953
!"nate"	1.0135135135135136	!"Tides"	1.0294117647058822	!"formulated"	1.0377358490566038	#"59"	1.1258741258741258
!"pc"	1.4404761904761905	!"Tonton"	1.2619047619047619	!"granulomas"	1.1774193548387097	#"7300"	1.037037037037037
!"pits"	1.05	!"Tortures"	1.04444444444444445	!"harles"	1.05	#"AF"	1.4666666666666666
		!"Ubiquitous"	1.0	!"hein"	1.0961538461538463	"# Data"	1.2066571772241242