

Submission Deadline: 31/05/2023 - 23:59

Question: 01

MAP REDUCE - DATASET: write map reduce program(s) using python-mr in colab notebook and complete the following exercises

- Find the files which contain the following phrases
 - "draping the folds of her velvet dress"
 - "grim clasp to the iron stanchions"
 - "Lo, I teach you the Superman!"
 - "Through that divine allegiance"
- Find the file which has maximum number of tokens
- Find the file which has the longest sentence (a sentence ends in a full stop and can span multiple lines)

Dataset:

- [pg1513.txt](#)
- [2160-0.txt](#)
- [pg11.txt](#)
- [pg2600.txt](#)
- [345.txt.utf-8](#)
- [64317.txt.utf-8](#)
- [70815.txt.utf-8](#)
- [1080-0.txt](#)
- [70817.txt.utf-8](#)
- [1998.txt.utf-8](#)

```
In [ ]: # # First make sure the directory you are supposed to download to, is available and acce
!mkdir -p dataset
!wget -P dataset https://www.gutenberg.org/cache/epub/1513/pg1513.txt
!wget -P dataset https://www.gutenberg.org/files/2160/2160-0.txt
!wget -P dataset https://www.gutenberg.org/cache/epub/11/pg11.txt
!wget -P dataset https://www.gutenberg.org/cache/epub/2600/pg2600.txt
!wget -P dataset https://www.gutenberg.org/ebooks/345.txt.utf-8
!wget -P dataset https://www.gutenberg.org/ebooks/64317.txt.utf-8
!wget -P dataset https://www.gutenberg.org/ebooks/70815.txt.utf-8
!wget -P dataset https://www.gutenberg.org/files/1080/1080-0.txt
!wget -P dataset https://www.gutenberg.org/ebooks/70817.txt.utf-8
!wget -P dataset https://www.gutenberg.org/ebooks/1998.txt.utf-8
```

```
--2023-06-01 20:54:57-- https://www.gutenberg.org/cache/epub/1513/pg1513.txt
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 169260 (165K) [text/plain]
Saving to: 'dataset/pg1513.txt'
```

```
pg1513.txt          100%[=====>] 165.29K  --.-KB/s    in 0.05s

2023-06-01 20:54:58 (3.38 MB/s) - 'dataset/pg1513.txt' saved [169260/169260]

--2023-06-01 20:54:58-- https://www.gutenberg.org/files/2160/2160-0.txt
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 887402 (867K) [text/plain]
Saving to: 'dataset/2160-0.txt'

2160-0.txt          100%[=====>] 866.60K  --.-KB/s    in 0.1s

2023-06-01 20:54:58 (8.88 MB/s) - 'dataset/2160-0.txt' saved [887402/887402]

--2023-06-01 20:54:58-- https://www.gutenberg.org/cache/epub/11/pg11.txt
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 174280 (170K) [text/plain]
Saving to: 'dataset/pg11.txt'

pg11.txt            100%[=====>] 170.20K  --.-KB/s    in 0.05s

2023-06-01 20:54:58 (3.48 MB/s) - 'dataset/pg11.txt' saved [174280/174280]

--2023-06-01 20:54:58-- https://www.gutenberg.org/cache/epub/2600/pg2600.txt
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 3359372 (3.2M) [text/plain]
Saving to: 'dataset/pg2600.txt'

pg2600.txt          100%[=====>]   3.20M  19.1MB/s    in 0.2s

2023-06-01 20:54:59 (19.1 MB/s) - 'dataset/pg2600.txt' saved [3359372/3359372]

--2023-06-01 20:54:59-- https://www.gutenberg.org/ebooks/345.txt.utf-8
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://www.gutenberg.org/cache/epub/345/pg345.txt [following]
--2023-06-01 20:54:59-- https://www.gutenberg.org/cache/epub/345/pg345.txt
Reusing existing connection to www.gutenberg.org:443.
HTTP request sent, awaiting response... 200 OK
Length: 881691 (861K) [text/plain]
Saving to: 'dataset/345.txt.utf-8'

345.txt.utf-8       100%[=====>] 861.03K  --.-KB/s    in 0.1s

2023-06-01 20:54:59 (6.97 MB/s) - 'dataset/345.txt.utf-8' saved [881691/881691]

--2023-06-01 20:54:59-- https://www.gutenberg.org/ebooks/64317.txt.utf-8
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://www.gutenberg.org/cache/epub/64317/pg64317.txt [following]
--2023-06-01 20:54:59-- https://www.gutenberg.org/cache/epub/64317/pg64317.txt
Reusing existing connection to www.gutenberg.org:443.
HTTP request sent, awaiting response... 200 OK
```

Length: 306227 (299K) [text/plain]
Saving to: 'dataset/64317.txt.utf-8'

64317.txt.utf-8 100%[=====>] 299.05K --.-KB/s in 0.09s

2023-06-01 20:54:59 (3.13 MB/s) - 'dataset/64317.txt.utf-8' saved [306227/306227]

--2023-06-01 20:54:59-- https://www.gutenberg.org/ebooks/70815.txt.utf-8
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://www.gutenberg.org/cache/epub/70815/pg70815.txt [following]
--2023-06-01 20:54:59-- https://www.gutenberg.org/cache/epub/70815/pg70815.txt
Reusing existing connection to www.gutenberg.org:443.
HTTP request sent, awaiting response... 200 OK
Length: 474995 (464K) [text/plain]
Saving to: 'dataset/70815.txt.utf-8'

70815.txt.utf-8 100%[=====>] 463.86K --.-KB/s in 0.1s

2023-06-01 20:55:00 (3.85 MB/s) - 'dataset/70815.txt.utf-8' saved [474995/474995]

--2023-06-01 20:55:00-- https://www.gutenberg.org/files/1080/1080-0.txt
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 39819 (39K) [text/plain]
Saving to: 'dataset/1080-0.txt'

1080-0.txt 100%[=====>] 38.89K --.-KB/s in 0.02s

2023-06-01 20:55:00 (1.57 MB/s) - 'dataset/1080-0.txt' saved [39819/39819]

--2023-06-01 20:55:00-- https://www.gutenberg.org/ebooks/70817.txt.utf-8
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://www.gutenberg.org/cache/epub/70817/pg70817.txt [following]
--2023-06-01 20:55:00-- https://www.gutenberg.org/cache/epub/70817/pg70817.txt
Reusing existing connection to www.gutenberg.org:443.
HTTP request sent, awaiting response... 200 OK
Length: 166627 (163K) [text/plain]
Saving to: 'dataset/70817.txt.utf-8'

70817.txt.utf-8 100%[=====>] 162.72K --.-KB/s in 0.07s

2023-06-01 20:55:00 (2.23 MB/s) - 'dataset/70817.txt.utf-8' saved [166627/166627]

--2023-06-01 20:55:00-- https://www.gutenberg.org/ebooks/1998.txt.utf-8
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 302 Found
Location: https://www.gutenberg.org/cache/epub/1998/pg1998.txt [following]
--2023-06-01 20:55:00-- https://www.gutenberg.org/cache/epub/1998/pg1998.txt
Reusing existing connection to www.gutenberg.org:443.
HTTP request sent, awaiting response... 200 OK
Length: 683036 (667K) [text/plain]
Saving to: 'dataset/1998.txt.utf-8'

1998.txt.utf-8 100%[=====>] 667.03K --.-KB/s in 0.1s

1. a) Find the files which contain the following phrases

- "draping the folds of her velvet dress"
- "grim clasp to the iron stanchions"
- "Lo, I teach you the Superman!"
- "Through that divine allegiance"

```
In [ ]: !pip install mrjob #installing mrjob
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting mrjob
  Downloading mrjob-0.7.4-py2.py3-none-any.whl (439 kB)
    _____ 439.6/439.6 kB 8.
0 MB/s eta 0:00:00
Requirement already satisfied: PyYAML>=3.10 in /usr/local/lib/python3.10/dist-packages (from mrjob) (6.0)
Installing collected packages: mrjob
Successfully installed mrjob-0.7.4
```

```
In [ ]: %%writefile Finder.py
```

```
from mrjob.job import MRJob
from mrjob.step import MRStep
import os
import re

class MRFinder(MRJob):
    count = 0

    def mapper(self, _, line):
        filename = os.getenv('mapreduce_map_input_file') # Get the filename
        phrases = ["draping the folds of her velvet dress",
                  "grim clasp to the iron stanchions",
                  "Lo, I teach you the Superman!",
                  "Through that divine allegiance"]

        for phrase in phrases:
            if phrase in line:
                yield filename, phrase

    def reducer(self, key, values):
        yield key, list(values)

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

Writing Finder.py

```
In [ ]: !python Finder.py /content/dataset/*
```

```
No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/Finder.root.20230601.174203.407862
Running step 1 of 1...
job output is in /tmp/Finder.root.20230601.174203.407862/output
Streaming final output from /tmp/Finder.root.20230601.174203.407862/output...
"file:///content/dataset/pg2600.txt" ["draping the folds of her velvet dress"]
"file:///content/dataset/1998.txt.utf-8" ["Lo, I teach you the Superman!"]
"file:///content/dataset/345.txt.utf-8" ["grim clasp to the iron stanchions"]
Removing temp directory /tmp/Finder.root.20230601.174203.407862...
```

So, I am getting the following sentences in the following files accordingly "file:///content/dataset/pg2600.txt" ["draping the folds of her velvet dress"] "file:///content/dataset/1998.txt.utf-8" ["Lo, I teach you the Superman!"] "file:///content/dataset/345.txt.utf-8" ["grim clasp to the iron stanchions"]

1.b) Find the file which has maximum number of tokens

```
In [ ]: %%writefile token_count.py

from mrjob.job import MRJob
from mrjob.step import MRStep
import os

class MRTokenCount(MRJob):

    def steps(self):
        return [
            MRStep(mapper=self.mapper,
                  reducer=self.reducer)
        ]

    def mapper(self, _, line):
        # send each line in the file with the filename to the reducer
        yield os.environ['mapreduce_map_input_file'], line

    def reducer(self, filename, lines):
        # calculate the total number of tokens in each file
        token_count = sum(len(line.split()) for line in lines)
        yield filename, token_count

if __name__ == "__main__":
    MRTokenCount.run()
```

Writing token_count.py

```
In [ ]: !python token_count.py /content/dataset/*

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/token_count.root.20230601.174249.146210
Running step 1 of 1...
job output is in /tmp/token_count.root.20230601.174249.146210/output
Streaming final output from /tmp/token_count.root.20230601.174249.146210/output...
"file:///content/dataset/1080-0.txt"      6512
"file:///content/dataset/1998.txt.utf-8" 113997
"file:///content/dataset/2160-0.txt"    151379
"file:///content/dataset/345.txt.utf-8" 164459
"file:///content/dataset/64317.txt.utf-8" 51257
"file:///content/dataset/70815.txt.utf-8" 82782
"file:///content/dataset/70817.txt.utf-8" 28184
"file:///content/dataset/pg11.txt"      29590
"file:///content/dataset/pg1513.txt"    29002
"file:///content/dataset/pg2600.txt"    566330
Removing temp directory /tmp/token_count.root.20230601.174249.146210...
```

So, from the above output I am getting file name pg2600.txt is having maximum number of tokens

1.c) Find the file which has the longest sentence (a sentence ends in a full stop and can span multiple lines)

```
In [ ]: %%writefile longest_sentence.py

from mrjob.job import MRJob
from mrjob.step import MRStep
import os
```

```

import re

class MRLongestSentence(MRJob):

    def steps(self):
        return [
            MRStep(mapper=self.mapper,
                    reducer=self.reducer),
            MRStep(reducer=self.reducer_find_max)
        ]

    def mapper(self, _, line):
        # split the line into sentences
        sentences = re.split('\.|\?|\!', line)
        for sentence in sentences:
            # send each sentence in the file with the filename to the reducer
            yield os.environ['mapreduce_map_input_file'], sentence.strip()

    def reducer(self, filename, sentences):
        # calculate the length of the longest sentence in each file
        longest_sentence = max(sentences, key=len)
        yield None, (len(longest_sentence), filename)

    def reducer_find_max(self, _, file_data):
        # find the file with the longest sentence
        yield max(file_data)

if __name__ == "__main__":
    MRLongestSentence.run()

```

Writing longest_sentence.py

```
In [ ]: !python longest_sentence.py /content/dataset/*
```

```

No configs found; falling back on auto-configuration
No configs specified for inline runner
Creating temp directory /tmp/longest_sentence.root.20230601.174328.333751
Running step 1 of 2...
Running step 2 of 2...
job output is in /tmp/longest_sentence.root.20230601.174328.333751/output
Streaming final output from /tmp/longest_sentence.root.20230601.174328.333751/output...
89      "file:///content/dataset/pg1513.txt"
Removing temp directory /tmp/longest_sentence.root.20230601.174328.333751...

```

So, I have got pg1513.txt file has the longest sentence. I have defined four methods. steps: define the steps of the MapReduce job. first step - a mapper and a reducer, and the second step includes only a reducer. Mapper method splits each line into sentences using re.split(). The reducer method takes a key-value pair, finds the longest sentence, and yields a new key-value pair. Reducer_find_max method takes a key-value pair and yields the maximum tuple. Finally, MapReduce job is executed with MRLongestSentence.run() in main function.

Question: 02

Write a PySpark RDD program to compute:

- The list of common (overlapping / intersection) words across all the ten files listed above.
- List the word frequency of common words across files as a dict (tip: use join method in the rdd). The size of the dict equals the list of common words and each key in the top level dict contains the word count corresponding to all the ten files

```
■ {"word":{"pg1513.txt":10,"2160-0.txt":25,...}}
```

```
In [ ]: !pip install pyspark #Installing pyspark
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting pyspark
  Downloading pyspark-3.4.0.tar.gz (310.8 MB)
    _____ 310.8/310.8 MB 2.5 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: py4j==0.10.9.7 in /usr/local/lib/python3.10/dist-packages (from pyspark) (0.10.9.7)
Building wheels for collected packages: pyspark
  Building wheel for pyspark (setup.py) ... done
  Created wheel for pyspark: filename=pyspark-3.4.0-py2.py3-none-any.whl size=311317130 sha256=d7a33d72b1aac218bc55f682989f8cfd32a39346bed1c7fa019a450593ca8348
  Stored in directory: /root/.cache/pip/wheels/7b/1b/4b/3363a1d04368e7ff0d408e57ff57966fcdf00583774e761327
Successfully built pyspark
Installing collected packages: pyspark
Successfully installed pyspark-3.4.0
```

```
In [ ]: from pyspark import SparkContext, SparkConf
import re
import os

# Create Spark Context
conf = SparkConf().setAppName("CommonWords")
sc = SparkContext.getOrCreate(conf=conf)

dir = os.listdir('/content/dataset')
files = [f"/content/dataset/{filename}" for filename in dir]

common_words_rdd = None

# Create RDD for each file, extract words, and perform union operation
for file in files:
    rdd = sc.textFile(file)
    words_rdd = rdd.flatMap(lambda line: re.findall(r'\b[a-zA-Z][a-zA-Z0-9]*(?:[.]*[a-zA-Z0-9])?', line))
    if common_words_rdd is None:
        common_words_rdd = words_rdd
    else:
        common_words_rdd = common_words_rdd.intersection(words_rdd)

# Collect the common words
common_words = common_words_rdd.collect()
print("Common Words are:", common_words)
print("Length of Common Words:", len(common_words))

# Create a dictionary to hold word frequencies
word_frequencies = {}

for file in files:
    rdd = sc.textFile(file)
    words_rdd = rdd.flatMap(lambda line: re.findall(r'\b[a-zA-Z][a-zA-Z0-9]*(?:[.]*[a-zA-Z0-9])?', line))
    word_count = words_rdd.countByValue()
    for word, count in word_count.items():
        if word not in word_frequencies:
            word_frequencies[word] = {}
        word_frequencies[word][file] = count

print('List the word frequency of common words across files: ', word_frequencies)
```

```
Common Words are: ['family', 'references', 'immediate', 'of', 'rather', 'two', 'away',
```

'right', 'official', 'something', 'gross', 'agent', 'was', 'put', 'think', 'would', 'putting', 'us', 'leave', 'agreement', 'particular', 'ein', 'appears', 'demand', 'thousand', 'equipment', 'manager', 'send', 'indicate', 'free', 'head', 'archive', 'earth', 'freely', 'an', 'requirements', 'lieu', 'sent', 'computers', 'especially', 'files', 'format', 'entity', 'expense', 'have', 'no', 'opportunity', 'till', 'making', 'derive', 'indirect', 'looked', 'glad', 'bound', 'and', 'follow', 'gone', 'warranty', 'federal', 'preserve', 'harmless', 'editions', 'compilation', 'distribute', 'ut', 'processing', 'indemnity', 'imposed', 'almost', 'which', 'most', 'displaying', 'required', 'you', 'by', 'reports', 'pg', 'ebooks', 'concerning', 'return', 'treatment', 'email', 'how', 'manner', 'codes', 'state', 'cease', 'govern', 'replace', 'several', 'checks', 'created', 'internal', 'version', 'myself', 'longer', 'these', 'ten', 'month', 'way', 'medium', 'donate', 'we', 'holder', 'print', 'negligence', 'array', 'change', 'this', 'rest', 'corrupt', 'based', 'project', 'compressed', 'punitive', 'prominently', 'search', 'modified', 'use', 'claim', 'abide', 'hart', 'gratefully', 'appearing', 'violates', 'new', 'form', 'incomplete', 'financial', 'title', 'things', 'inaccurate', 'consequential', 'meet', 'pglaf', 'extent', 'visit', 'whose', 'carry', 'fell', 'data', 'sentence', 'fast', 'its', 'body', 'under', 'exporting', 'previous', 'found', 'small', 'about', 'original', 'friend', 'also', 'contain', 'states', 'north', 'updated', 'it', 'not', 'reading', 'since', 'merchantability', 'maximum', 'my', 'profits', 'concept', 'defective', 'so', 'applicable', 'possessed', 'constant', 'immediately', 'or', 'discontinue', 'easy', 'get', 'f', 'contact', 'alternative', 'direct', 's', 'common', 'there', 'turn', 'even', 'redistribution', 'copied', 'stored', 'accessed', 'charges', 'i', 'last', 'alteration', 'salt', 'came', 'pay', 'are', 'offers', 'machine', 'set', 'matter', 'remaining', 'different', 'disclaimers', 'forty', 'check', 'paragraph', 'identify', 'sharing', 'charge', 'modification', 'produced', 'deletions', 'mean', 'depends', 'comply', 'measure', 'trouble', 'confirmation', 'may', 'good', 'houses', 'b', 'provide', 'now', 'formats', 'men', 'number', 'office', 'license', 'off', 'donations', 'organized', 'elect', 'generations', 'purpose', 'performing', 'word', 'makes', 'hold', 'nothing', 'nearly', 'among', 'to', 'credit', 'require', 're', 'directly', 'granted', 'calculated', 'creation', 'foundation', 'learn', 'their', 'written', 'whenever', 'table', 'has', 'secure', 'non', 'lot', 'fitness', 'cost', 'invalidity', 'protect', 'people', 'such', 'viewed', 'should', 'revenue', 'opportunities', 'twelve', 'proofread', 'sure', 'beat', 'calculate', 'occur', 'able', 'home', 'disk', 'writing', 'stand', 'loose', 'cause', 'edition', 'limited', 'child', 'despite', 'efforts', 'move', 'creating', 'figure', 'liability', 'ascii', 'linked', 'donors', 'certain', 'kind', 'three', 'accepted', 'group', 'contract', 'house', 'expenses', 'restrictions', 'read', 'anyone', 'am', 'exists', 'take', 'u.s', 'prevent', 'collection', 'your', 'laws', 'present', 'includes', 'future', 'swamp', 'will', 'could', 'unless', 'times', 'often', 'explanation', 'practically', 'electronic', 'some', 'children', 'file', 'another', 'for', 'permission', 'donation', 'they', 'girl', 'critical', 'being', 'increasing', 'provided', 'been', 'unsolicited', 'performances', 'half', 'irs', 'made', 'same', 'survive', 'hope', 'royalty', 'find', 'included', 'deductible', 'must', 'additional', 'ever', 'sound', 'current', 'believe', 'research', 'where', 'hot', 'remedies', 'indirectly', 'business', 'downloading', 'used', 'interpreted', 'network', 'breach', 'worse', 'possibility', 'promotion', 'below', 'known', 'exempt', 'them', 'give', 'binary', 'remove', 'only', 'electronically', 'name', 'arise', 'perform', 'best', 'marked', 'man', 'arms', 'unlink', 'many', 'access', 'request', 'addition', 'grown', 'paper', 'doubt', 'terms', 'a', 'solicitation', 'keep', 'who', 'notifies', 'giving', 'works', 'shared', 'old', 'outdated', 'costs', 'stay', 'trademark', 'from', 'part', 'either', 'defect', 'thing', 'having', 'detach', 'information', 'hundreds', 'damages', 'derivative', 'contributions', 'paragraphs', 'need', 'through', 'attached', 'containing', 'talk', 'receive', 'party', 'accessible', 'ago', 'before', 'ready', 'registered', 'help', 'true', 'anything', 'never', 'die', 'four', 'support', 'bring', 'though', 'charities', 'is', 'strength', 'bear', 'synonymous', 'share', 'viewing', 'rules', 'phrase', 'obtain', 'displayed', 'volunteer', 'at', 'tax', 'charitable', 'said', 'following', 'mission', 'permanent', 'than', 'law', 'united', 'volunteers', 'damage', 'her', 'allow', 'language', 'links', 'www.gutenberg.org', 'thus', 'renamed', 'prohibition', 'solicit', 'poor', 'refund', 'compliance', 'widest', 'countries', 'copy', 'goals', 'alone', 'copies', 'strict', 'five', 'accordance', 'knew', 'any', 'available', 'received', 'soon', 'taxes', 'web', 'paperwork', 'remain', 'offer', 'disclaimer', 'english', 'limitation', 'educational', 'care', 'within', 'parts', 'necessarily', 'much', 'were', 'attempt', 'whatsoever', 'both', 'days', 'heart', 'middle', 'cannot', 'derived', 'returns', 'additions', 'proprietary', 'walk', 'produce', 'far', 'already', 'service', 'against', 'card', 'expend', 'out', 'twenty', 'in', 'provision', 'nor', 'long', 'production', 'subscribe', 'include', 'actual', 'brought', 'pages', 'physical', 'identification', 'shall', 'e', 'city', 'variety', 'be', 'first', 'posted', 'forth', 'periodic', 'agreed', 'without', 'hype', 'text', 'except', 'destroy', 'few', 'every', 'wish', 'methods', 'outside', 'facility', 'whom', 'general', 'wife', 'employees', 'main', 'apply', 'notice', 'accept', 'seem', 'se

ction', 'intellectual', 'anywhere', 'perhaps', 'further', 'told', 'online', 'own', 'more', 'addresses', 'just', 'regulating', 'past', 'fight', 'after', 'removed', 'person', 'make', 'easily', 'legal', 'providing', 'public', 'upon', 'discover', 'ebook', 'confirmed', 'fine', 'ask', 'redistribute', 'implied', 'very', 'appear', 'transcribe', 'world', 'neither', 'paid', 'release', 'distributing', 'one', 'located', 'all', 'met', 'takes', 'express', 'come', 'distributed', 'because', 'choose', 'west', 'while', 'literary', 'virus', 'effort', 'problem', 'grow', 'want', 'disclaim', 'readable', 'receiving', 'obtaining', 'warranties', 'place', 'second', 'on', 'hundred', 'if', 'agree', 'does', 'newsletter', 'status', 'bent', 'less', 'answer', 'alive', 'see', 'lake', 'plain', 'me', 'associated', 'reaching', 'owner', 'complying', 'staff', 'when', 'copyright', 'work', 'active', 'c', 'page', 'he', 'library', 'above', 'obsolete', 'again', 'however', 'subject', 'accepting', 'like', 'prepare', 'promoting', 'other', 'vanilla', 'others', 'country', 'yet', 'michael', 'errors', 'copying', 'start', 'gutenberg', 'let', 'particularly', 'owns', 'transcription', 'maintaining', 'approach', 'distribution', 'but', 'years', 'do', 'money', 'commercial', 'using', 'unenforceability', 'young', 'individual', 'licensed', 'indicating', 'property', 'method', 'specified', 'provisions', 'ensuring', 'doors', 'committed', 'clearly', 'determine', 'considerable', 'sending', 'employee', 'whole', 'replacement', 'date', 'author', 'taken', 'distributor', 'uniform', 'up', 'reason', 'each', 'age', 'special', 'types', 'might', 'payments', 'that', 'life', 'little', 'please', 'the', 'legally', 'user', 'beginning', 'death', 'receipt', 'international', 'sort', 'liable', 'order', 'great', 'then', 'redistributing', 'time', 'him', 'address', 'full', 'know', 'assistance', 'walks', 'domain', 'indemnify', 'fee', 'representations', 'profit', 'indeed', 'incidental', 'nonproprietary', 'understand', 'fees', 'paying', 'sections', 'fix', 'late', 'performed', 'mississippi', 'his', 'hear', 'possession', 'means', 'as', 'locations', 'into', 'd', 'effect', 'aged', 'reported', 'cold', 'end', 'convert', 'permitted', 'discovered', 'too', 'exclusion', 'display', 'important', 'those', 'infringement', 'keeping', 'owed', 'including', 'better', 'royalties', 'fair', 'printed', 'computer', 'day', 'well', 'what', 'void', 'with', 'course', 'defects', 'originator', 'our', 'six', 'reasonable', 'corporation', 'least', 'can', 'ways', 'statements', 'she', 'described', 'professor', 'given', 'damaged']

Length of Common Words: 819

List the word frequency of common words across files: {'the': {'/content/dataset/1998.txt.utf-8': 6132, '/content/dataset/pg1513.txt': 877, '/content/dataset/pg11.txt': 1839, '/content/dataset/2160-0.txt': 8667, '/content/dataset/345.txt.utf-8': 8104, '/content/dataset/64317.txt.utf-8': 2594, '/content/dataset/70817.txt.utf-8': 1914, '/content/dataset/1080-0.txt': 359, '/content/dataset/pg2600.txt': 34737, '/content/dataset/70815.txt.utf-8': 6213}, 'project': {'/content/dataset/1998.txt.utf-8': 88, '/content/dataset/pg1513.txt': 89, '/content/dataset/pg11.txt': 88, '/content/dataset/2160-0.txt': 94, '/content/dataset/345.txt.utf-8': 89, '/content/dataset/64317.txt.utf-8': 90, '/content/dataset/70817.txt.utf-8': 89, '/content/dataset/1080-0.txt': 89, '/content/dataset/pg2600.txt': 101, '/content/dataset/70815.txt.utf-8': 90}, 'gutenberg': {'/content/dataset/1998.txt.utf-8': 88, '/content/dataset/pg1513.txt': 89, '/content/dataset/pg11.txt': 88, '/content/dataset/2160-0.txt': 87, '/content/dataset/345.txt.utf-8': 88, '/content/dataset/64317.txt.utf-8': 89, '/content/dataset/70817.txt.utf-8': 88, '/content/dataset/1080-0.txt': 88, '/content/dataset/pg2600.txt': 88, '/content/dataset/70815.txt.utf-8': 88}, 'ebook': {'/content/dataset/1998.txt.utf-8': 13, '/content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 10, '/content/dataset/345.txt.utf-8': 13, '/content/dataset/64317.txt.utf-8': 13, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 13, '/content/dataset/pg2600.txt': 13, '/content/dataset/70815.txt.utf-8': 13}, 'of': {'/content/dataset/1998.txt.utf-8': 2911, '/content/dataset/pg1513.txt': 518, '/content/dataset/pg11.txt': 638, '/content/dataset/2160-0.txt': 5682, '/content/dataset/345.txt.utf-8': 3756, '/content/dataset/64317.txt.utf-8': 1242, '/content/dataset/70817.txt.utf-8': 770, '/content/dataset/1080-0.txt': 261, '/content/dataset/pg2600.txt': 15013, '/content/dataset/70815.txt.utf-8': 3108}, 'thus': {'/content/dataset/1998.txt.utf-8': 422, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 27, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 27, '/content/dataset/70815.txt.utf-8': 37}, 'by': {'/content/dataset/1998.txt.utf-8': 328, '/content/dataset/pg1513.txt': 145, '/content/dataset/pg11.txt': 86, '/content/dataset/2160-0.txt': 969, '/content/dataset/345.txt.utf-8': 546, '/content/dataset/64317.txt.utf-8': 150, '/content/dataset/70817.txt.utf-8': 105, '/content/dataset/1080-0.txt': 61, '/content/dataset/pg2600.txt': 2465, '/content/dataset/70815.txt.utf-8': 382}, 'this': {'/content/dataset/1998.txt.utf-8': 682, '/content/dataset/pg1513.txt': 266, '/content/dataset/pg11.txt': 178, '/content/dataset/2160-0.txt': 1014, '/content/dataset/345.txt.utf-8': 676, '/content/dataset/64317.txt.utf-8': 212, '/content/dataset/70817.txt.utf-8': 85, '/content/dataset/1080-0.txt': 73, '/co

content/dataset/pg2600.txt': 2120, '/content/dataset/70815.txt': 314}, 'is': {'content/dataset/1998.txt.utf-8': 1841, '/content/dataset/pg1513.txt': 372, '/content/dataset/pg11.txt': 123, '/content/dataset/2160-0.txt': 1325, '/content/dataset/345.txt.utf-8': 1525, '/content/dataset/64317.txt.utf-8': 128, '/content/dataset/70817.txt.utf-8': 61, '/content/dataset/1080-0.txt': 44, '/content/dataset/pg2600.txt': 3344, '/content/dataset/70815.txt.utf-8': 405}, 'for': {'content/dataset/1998.txt.utf-8': 1169, '/content/dataset/pg1513.txt': 251, '/content/dataset/pg11.txt': 179, '/content/dataset/2160-0.txt': 1277, '/content/dataset/345.txt.utf-8': 1564, '/content/dataset/64317.txt.utf-8': 339, '/content/dataset/70817.txt.utf-8': 170, '/content/dataset/1080-0.txt': 74, '/content/dataset/pg2600.txt': 3555, '/content/dataset/70815.txt.utf-8': 607}, 'use': {'content/dataset/1998.txt.utf-8': 23, '/content/dataset/pg1513.txt': 28, '/content/dataset/pg11.txt': 32, '/content/dataset/2160-0.txt': 72, '/content/dataset/345.txt.utf-8': 56, '/content/dataset/64317.txt.utf-8': 22, '/content/dataset/70817.txt.utf-8': 15, '/content/dataset/1080-0.txt': 15, '/content/dataset/pg2600.txt': 86, '/content/dataset/70815.txt.utf-8': 25}, 'anyone': {'content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 15, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 193, '/content/dataset/70815.txt.utf-8': 16}, 'anywhere': {'content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 26, '/content/dataset/70815.txt.utf-8': 4}, 'in': {'content/dataset/1998.txt.utf-8': 1557, '/content/dataset/pg1513.txt': 394, '/content/dataset/pg11.txt': 434, '/content/dataset/2160-0.txt': 3224, '/content/dataset/345.txt.utf-8': 2571, '/content/dataset/64317.txt.utf-8': 867, '/content/dataset/70817.txt.utf-8': 469, '/content/dataset/1080-0.txt': 132, '/content/dataset/pg2600.txt': 9039, '/content/dataset/70815.txt.utf-8': 1169}, 'united': {'content/dataset/1998.txt.utf-8': 15, '/content/dataset/pg1513.txt': 15, '/content/dataset/pg11.txt': 15, '/content/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 15, '/content/dataset/70817.txt.utf-8': 15, '/content/dataset/1080-0.txt': 15, '/content/dataset/pg2600.txt': 31, '/content/dataset/70815.txt.utf-8': 16}, 'states': {'content/dataset/1998.txt.utf-8': 20, '/content/dataset/pg1513.txt': 19, '/content/dataset/pg11.txt': 19, '/content/dataset/2160-0.txt': 15, '/content/dataset/345.txt.utf-8': 22, '/content/dataset/64317.txt.utf-8': 20, '/content/dataset/70817.txt.utf-8': 19, '/content/dataset/1080-0.txt': 19, '/content/dataset/pg2600.txt': 25, '/content/dataset/70815.txt.utf-8': 19}, 'and': {'content/dataset/1998.txt.utf-8': 4700, '/content/dataset/pg1513.txt': 806, '/content/dataset/pg11.txt': 941, '/content/dataset/2160-0.txt': 5077, '/content/dataset/345.txt.utf-8': 5974, '/content/dataset/64317.txt.utf-8': 1635, '/content/dataset/70817.txt.utf-8': 797, '/content/dataset/1080-0.txt': 184, '/content/dataset/pg2600.txt': 22303, '/content/dataset/70815.txt.utf-8': 2553}, 'most': {'content/dataset/1998.txt.utf-8': 152, '/content/dataset/pg1513.txt': 22, '/content/dataset/pg11.txt': 14, '/content/dataset/2160-0.txt': 177, '/content/dataset/345.txt.utf-8': 81, '/content/dataset/64317.txt.utf-8': 37, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 14, '/content/dataset/pg2600.txt': 330, '/content/dataset/70815.txt.utf-8': 37}, 'other': {'content/dataset/1998.txt.utf-8': 117, '/content/dataset/pg1513.txt': 34, '/content/dataset/pg11.txt': 57, '/content/dataset/2160-0.txt': 276, '/content/dataset/345.txt.utf-8': 191, '/content/dataset/64317.txt.utf-8': 82, '/content/dataset/70817.txt.utf-8': 34, '/content/dataset/1080-0.txt': 26, '/content/dataset/pg2600.txt': 598, '/content/dataset/70815.txt.utf-8': 158}, 'parts': {'content/dataset/1998.txt.utf-8': 7, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 39, '/content/dataset/345.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 24, '/content/dataset/70815.txt.utf-8': 2}, 'world': {'content/dataset/1998.txt.utf-8': 159, '/content/dataset/pg1513.txt': 21, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 48, '/content/dataset/345.txt.utf-8': 64, '/content/dataset/64317.txt.utf-8': 35, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 167, '/content/dataset/70815.txt.utf-8': 93}, 'at': {'content/dataset/1998.txt.utf-8': 514, '/content/dataset/pg1513.txt': 83, '/content/dataset/pg11.txt': 224, '/content/dataset/2160-0.txt': 1120, '/content/dataset/345.txt.utf-8': 1098, '/content/dataset/64317.txt.utf-8': 421, '/content/dataset/70817.txt.utf-8': 211, '/content/dataset/1080-0.txt': 42, '/content/dataset/pg2600.txt': 4545, '/content/dataset/70815.txt.utf-8': 408}, 'no': {'content/dataset/1998.txt.utf-8': 283, '/content/dataset/pg1513.txt': 112, '/content/dataset/pg11.txt': 99, '/content/dataset/2160-0.txt': 480, '/content/dataset/345.txt.utf-8': 499, '/content/dataset/64317.txt.utf-8': 113, '/content/dataset/70817.txt.utf-8': 81, '/content/dataset/1080-0.txt': 22, '/content/dataset/pg2600.txt': 1411, '/content

nt/dataset/70815.txt.utf-8': 273}, 'cost': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 26, '/content/dataset/70815.txt.utf-8': 3}, 'with': {'/content/dataset/1998.txt.utf-8': 875, '/content/dataset/pg1513.txt': 300, '/content/dataset/pg11.txt': 227, '/content/dataset/2160-0.txt': 1711, '/content/dataset/345.txt.utf-8': 1334, '/content/dataset/64317.txt.utf-8': 471, '/content/dataset/70817.txt.utf-8': 223, '/content/dataset/1080-0.txt': 59, '/content/dataset/pg2600.txt': 5710, '/content/dataset/70815.txt.utf-8': 613}, 'almost': {'/content/dataset/1998.txt.utf-8': 29, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 31, '/content/dataset/345.txt.utf-8': 72, '/content/dataset/64317.txt.utf-8': 24, '/content/dataset/70817.txt.utf-8': 14, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 156, '/content/dataset/70815.txt.utf-8': 92}, 'restrictions': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 3, '/content/dataset/70815.txt.utf-8': 2}, 'whatsoever': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 9, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 2}, 'you': {'/content/dataset/1998.txt.utf-8': 373, '/content/dataset/pg1513.txt': 368, '/content/dataset/pg11.txt': 471, '/content/dataset/2160-0.txt': 816, '/content/dataset/345.txt.utf-8': 1486, '/content/dataset/64317.txt.utf-8': 597, '/content/dataset/70817.txt.utf-8': 148, '/content/dataset/1080-0.txt': 75, '/content/dataset/pg2600.txt': 3876, '/content/dataset/70815.txt.utf-8': 575}, 'may': {'/content/dataset/1998.txt.utf-8': 178, '/content/dataset/pg1513.txt': 64, '/content/dataset/pg11.txt': 28, '/content/dataset/2160-0.txt': 235, '/content/dataset/345.txt.utf-8': 431, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/70817.txt.utf-8': 25, '/content/dataset/1080-0.txt': 26, '/content/dataset/pg2600.txt': 292, '/content/dataset/70815.txt.utf-8': 87}, 'copy': {'/content/dataset/1998.txt.utf-8': 14, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 14, '/content/dataset/345.txt.utf-8': 21, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 12, '/content/dataset/1080-0.txt': 12, '/content/dataset/pg2600.txt': 16, '/content/dataset/70815.txt.utf-8': 13}, 'it': {'/content/dataset/1998.txt.utf-8': 1531, '/content/dataset/pg1513.txt': 244, '/content/dataset/pg11.txt': 607, '/content/dataset/2160-0.txt': 889, '/content/dataset/345.txt.utf-8': 2188, '/content/dataset/64317.txt.utf-8': 621, '/content/dataset/70817.txt.utf-8': 338, '/content/dataset/1080-0.txt': 41, '/content/dataset/pg2600.txt': 5617, '/content/dataset/70815.txt.utf-8': 915}, 'give': {'/content/dataset/1998.txt.utf-8': 75, '/content/dataset/pg1513.txt': 51, '/content/dataset/pg11.txt': 16, '/content/dataset/2160-0.txt': 126, '/content/dataset/345.txt.utf-8': 82, '/content/dataset/64317.txt.utf-8': 22, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 282, '/content/dataset/70815.txt.utf-8': 20}, 'away': {'/content/dataset/1998.txt.utf-8': 127, '/content/dataset/pg1513.txt': 26, '/content/dataset/pg11.txt': 28, '/content/dataset/2160-0.txt': 57, '/content/dataset/345.txt.utf-8': 192, '/content/dataset/64317.txt.utf-8': 78, '/content/dataset/70817.txt.utf-8': 36, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 617, '/content/dataset/70815.txt.utf-8': 72}, 'or': {'/content/dataset/1998.txt.utf-8': 294, '/content/dataset/pg1513.txt': 145, '/content/dataset/pg11.txt': 155, '/content/dataset/2160-0.txt': 489, '/content/dataset/345.txt.utf-8': 540, '/content/dataset/64317.txt.utf-8': 179, '/content/dataset/70817.txt.utf-8': 126, '/content/dataset/1080-0.txt': 113, '/content/dataset/pg2600.txt': 1602, '/content/dataset/70815.txt.utf-8': 297}, 're': {'/content/dataset/1998.txt.utf-8': 9, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 40, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 63, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 154, '/content/dataset/70815.txt.utf-8': 6}, 'under': {'/content/dataset/1998.txt.utf-8': 34, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 22, '/content/dataset/2160-0.txt': 113, '/content/dataset/345.txt.utf-8': 69, '/content/dataset/64317.txt.utf-8': 38, '/content/dataset/70817.txt.utf-8': 19, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 372, '/content/dataset/70815.txt.utf-8': 21}, 'terms': {'/content/dataset/1998.txt.utf-8': 25, '/content/dataset/pg1513.txt': 23, '/content/dataset/pg11.txt': 23, '/content/dataset/2160-0.txt': 41, '/content/dataset/345.txt.utf-8': 24, '/content/dataset/64317.txt.utf-8': 24, '/content/dataset/70817.txt.utf-8': 22, '/content/dataset/1080-0.txt': 22, '/content/dataset/pg2600.txt': 39, '/content/dataset/70815.txt.utf-8': 22}, 'license': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}

aset/1998.txt.utf-8': 18, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 18, '/content/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 18, '/content/dataset/pg2600.txt': 18, '/content/dataset/70815.txt.utf-8': 18}, 'included': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 11, '/content/dataset/70815.txt.utf-8': 6}, 'online': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 4, '/content/dataset/70815.txt.utf-8': 5}, 'www.gutenberg.org': {'/content/dataset/1998.txt.utf-8': 10, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 10, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 9, '/content/dataset/70815.txt.utf-8': 9}, 'if': {'/content/dataset/1998.txt.utf-8': 243, '/content/dataset/pg1513.txt': 106, '/content/dataset/pg11.txt': 118, '/content/dataset/2160-0.txt': 446, '/content/dataset/345.txt.utf-8': 660, '/content/dataset/64317.txt.utf-8': 168, '/content/dataset/70817.txt.utf-8': 67, '/content/dataset/1080-0.txt': 27, '/content/dataset/pg2600.txt': 1316, '/content/dataset/70815.txt.utf-8': 209}, 'are': {'/content/dataset/1998.txt.utf-8': 507, '/content/dataset/pg1513.txt': 89, '/content/dataset/pg11.txt': 73, '/content/dataset/2160-0.txt': 562, '/content/dataset/345.txt.utf-8': 607, '/content/dataset/64317.txt.utf-8': 69, '/content/dataset/70817.txt.utf-8': 33, '/content/dataset/1080-0.txt': 45, '/content/dataset/pg2600.txt': 1293, '/content/dataset/70815.txt.utf-8': 296}, 'not': {'/content/dataset/1998.txt.utf-8': 1018, '/content/dataset/pg1513.txt': 288, '/content/dataset/pg11.txt': 166, '/content/dataset/2160-0.txt': 930, '/content/dataset/345.txt.utf-8': 1430, '/content/dataset/64317.txt.utf-8': 101, '/content/dataset/70817.txt.utf-8': 116, '/content/dataset/1080-0.txt': 44, '/content/dataset/pg2600.txt': 4703, '/content/dataset/70815.txt.utf-8': 586}, 'located': {'/content/dataset/1998.txt.utf-8': 7, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 7, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 8, '/content/dataset/70815.txt.utf-8': 12}, 'will': {'/content/dataset/1998.txt.utf-8': 433, '/content/dataset/pg1513.txt': 157, '/content/dataset/pg11.txt': 39, '/content/dataset/2160-0.txt': 397, '/content/dataset/345.txt.utf-8': 463, '/content/dataset/64317.txt.utf-8': 29, '/content/dataset/70817.txt.utf-8': 22, '/content/dataset/1080-0.txt': 45, '/content/dataset/pg2600.txt': 1076, '/content/dataset/70815.txt.utf-8': 132}, 'have': {'/content/dataset/1998.txt.utf-8': 677, '/content/dataset/pg1513.txt': 133, '/content/dataset/pg11.txt': 83, '/content/dataset/2160-0.txt': 922, '/content/dataset/345.txt.utf-8': 1069, '/content/dataset/64317.txt.utf-8': 145, '/content/dataset/70817.txt.utf-8': 71, '/content/dataset/1080-0.txt': 36, '/content/dataset/pg2600.txt': 1982, '/content/dataset/70815.txt.utf-8': 406}, 'to': {'/content/dataset/1998.txt.utf-8': 2793, '/content/dataset/pg1513.txt': 626, '/content/dataset/pg11.txt': 811, '/content/dataset/2160-0.txt': 4443, '/content/dataset/345.txt.utf-8': 4748, '/content/dataset/64317.txt.utf-8': 1213, '/content/dataset/70817.txt.utf-8': 582, '/content/dataset/1080-0.txt': 190, '/content/dataset/pg2600.txt': 16759, '/content/dataset/70815.txt.utf-8': 2050}, 'check': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 11, '/content/dataset/70815.txt.utf-8': 5}, 'laws': {'/content/dataset/1998.txt.utf-8': 13, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 11, '/content/dataset/64317.txt.utf-8': 11, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 11, '/content/dataset/pg2600.txt': 99, '/content/dataset/70815.txt.utf-8': 14}, 'country': {'/content/dataset/1998.txt.utf-8': 11, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 163, '/content/dataset/345.txt.utf-8': 29, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 17, '/content/dataset/pg2600.txt': 110, '/content/dataset/70815.txt.utf-8': 70}, 'where': {'/content/dataset/1998.txt.utf-8': 145, '/content/dataset/pg1513.txt': 63, '/content/dataset/pg11.txt': 20, '/content/dataset/2160-0.txt': 224, '/content/dataset/345.txt.utf-8': 221, '/content/dataset/64317.txt.utf-8': 65, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 630, '/content/dataset/70815.txt.utf-8': 133}, 'before': {'/content/dataset/1998.txt.utf-8': 116, '/content/d

pg1513.txt': 24, '/content/dataset/pg11.txt': 41, '/content/dataset/2160-0.txt': 125, '/content/dataset/345.txt.utf-8': 281, '/content/dataset/64317.txt.utf-8': 94, '/content/dataset/70817.txt.utf-8': 61, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 835, '/content/dataset/70815.txt.utf-8': 121}, 'using': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 18, '/content/dataset/70815.txt.utf-8': 12}, 'title': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 11, '/content/dataset/345.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 2}, 'a': {'/content/dataset/1998.txt.utf-8': 1847, '/content/dataset/pg1513.txt': 547, '/content/dataset/pg11.txt': 695, '/content/dataset/2160-0.txt': 3744, '/content/dataset/345.txt.utf-8': 3013, '/content/dataset/64317.txt.utf-8': 1470, '/content/dataset/70817.txt.utf-8': 726, '/content/dataset/1080-0.txt': 153, '/content/dataset/pg2600.txt': 10583, '/content/dataset/70815.txt.utf-8': 1584}, 'all': {'/content/dataset/1998.txt.utf-8': 916, '/content/dataset/pg1513.txt': 114, '/content/dataset/pg11.txt': 197, '/content/dataset/2160-0.txt': 715, '/content/dataset/345.txt.utf-8': 1183, '/content/dataset/64317.txt.utf-8': 239, '/content/dataset/70817.txt.utf-8': 89, '/content/dataset/1080-0.txt': 27, '/content/dataset/pg2600.txt': 2812, '/content/dataset/70815.txt.utf-8': 187}, 'author': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 5}, 'common': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 44, '/content/dataset/345.txt.utf-8': 23, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 50, '/content/dataset/70815.txt.utf-8': 11}, 'release': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 11, '/content/dataset/70815.txt.utf-8': 3}, 'date': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 6, '/content/dataset/70815.txt.utf-8': 3}, 'updated': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 1}, 'language': {'/content/dataset/1998.txt.utf-8': 18, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 24, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 9, '/content/dataset/70815.txt.utf-8': 31}, 'english': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 7, '/content/dataset/2160-0.txt': 53, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 31, '/content/dataset/70815.txt.utf-8': 3}, 'produced': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 38, '/content/dataset/345.txt.utf-8': 10, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 42, '/content/dataset/70815.txt.utf-8': 2}, 'pg': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 10, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 's': {'/content/dataset/1998.txt.utf-8': 352, '/content/dataset/pg1513.txt': 317, '/content/dataset/pg11.txt': 211, '/content/dataset/2160-0.txt': 570, '/content/dataset/345.txt.utf-8': 736, '/content/dataset/64317.txt.utf-8': 480, '/content/dataset/70817.txt.utf-8': 97, '/content/dataset/1080-0.txt': 14, '/content/dataset/pg2600.txt': 4410, '/content/dataset/70815.txt.utf-8': 222}, 'beeen': {'/content/dataset/1998.txt.utf-8': 153, '/content/dataset/pg1513.txt': 20, '/content/dataset/pg11.txt': 38, '/content/dataset/2160-0.txt': 394, '/content/dataset/345.txt.u

tf-8': 393, '/content/dataset/64317.txt.utf-8': 107, '/content/dataset/70817.txt.utf-8': 68, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 1476, '/content/dataset/70815.txt.utf-8': 269}, 'particular': {'/content/dataset/1998.txt.utf-8': 7, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 66, '/content/dataset/345.txt.utf-8': 10, '/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 45, '/content/dataset/70815.txt.utf-8': 9}, 'often': {'/content/dataset/1998.txt.utf-8': 45, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 40, '/content/dataset/345.txt.utf-8': 19, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 140, '/content/dataset/70815.txt.utf-8': 23}, 'several': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 32, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 189, '/content/dataset/70815.txt.utf-8': 23}, 'paragraphs': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 5, '/content/dataset/70815.txt.utf-8': 3}, 'start': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 29, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 33, '/content/dataset/70815.txt.utf-8': 13}, 'first': {'/content/dataset/1998.txt.utf-8': 116, '/content/dataset/pg1513.txt': 34, '/content/dataset/pg11.txt': 51, '/content/dataset/2160-0.txt': 118, '/content/dataset/345.txt.utf-8': 135, '/content/dataset/64317.txt.utf-8': 56, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 633, '/content/dataset/70815.txt.utf-8': 93}, 'part': {'/content/dataset/1998.txt.utf-8': 52, '/content/dataset/pg1513.txt': 28, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 142, '/content/dataset/345.txt.utf-8': 63, '/content/dataset/64317.txt.utf-8': 13, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 237, '/content/dataset/70815.txt.utf-8': 23}, 'i': {'/content/dataset/1998.txt.utf-8': 1812, '/content/dataset/pg1513.txt': 659, '/content/dataset/pg11.txt': 524, '/content/dataset/2160-0.txt': 2790, '/content/dataset/345.txt.utf-8': 4846, '/content/dataset/64317.txt.utf-8': 1393, '/content/dataset/70817.txt.utf-8': 390, '/content/dataset/1080-0.txt': 55, '/content/dataset/pg2600.txt': 4540, '/content/dataset/70815.txt.utf-8': 543}, 'three': {'/content/dataset/1998.txt.utf-8': 33, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 28, '/content/dataset/2160-0.txt': 79, '/content/dataset/345.txt.utf-8': 61, '/content/dataset/64317.txt.utf-8': 38, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 277, '/content/dataset/70815.txt.utf-8': 76}, 'body': {'/content/dataset/1998.txt.utf-8': 72, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 44, '/content/dataset/345.txt.utf-8': 55, '/content/dataset/64317.txt.utf-8': 20, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 100, '/content/dataset/70815.txt.utf-8': 54}, 'reading': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 85, '/content/dataset/70815.txt.utf-8': 5}, 'writing': {'/content/dataset/1998.txt.utf-8': 11, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 13, '/content/dataset/345.txt.utf-8': 22, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 44, '/content/dataset/70815.txt.utf-8': 4}, 'on': {'/content/dataset/1998.txt.utf-8': 564, '/content/dataset/pg1513.txt': 94, '/content/dataset/pg11.txt': 204, '/content/dataset/2160-0.txt': 538, '/content/dataset/345.txt.utf-8': 1084, '/content/dataset/64317.txt.utf-8': 372, '/content/dataset/70817.txt.utf-8': 198, '/content/dataset/1080-0.txt': 19, '/content/dataset/pg2600.txt': 4014, '/content/dataset/70815.txt.utf-8': 151}, 'death': {'/content/dataset/1998.txt.utf-8': 73, '/content/dataset/pg1513.txt': 74, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 23, '/content/dataset/345.txt.utf-8': 95, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 24, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 191, '/content/dataset/70815.txt.utf-8': 34}, 'new': {'/content/dataset/1998.txt.utf-8': 157, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 70, '/content/dataset/345.txt.utf-8': 82, '/content/dataset/64317.txt.utf-8': 73, '/content/dataset/70817.txt.utf-

8': 24, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 290, '/content/dataset/70815.txt.utf-8': 47}, 'place': {'/content/dataset/1998.txt.utf-8': 62, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 155, '/content/dataset/345.txt.utf-8': 138, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/70817.txt.utf-8': 15, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 311, '/content/dataset/70815.txt.utf-8': 29}, 'friend': {'/content/dataset/1998.txt.utf-8': 60, '/content/dataset/pg1513.txt': 15, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 115, '/content/dataset/345.txt.utf-8': 186, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 186, '/content/dataset/70815.txt.utf-8': 30}, 'thousand': {'/content/dataset/1998.txt.utf-8': 23, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 52, '/content/dataset/345.txt.utf-8': 10, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 15, '/content/dataset/pg2600.txt': 171, '/content/dataset/70815.txt.utf-8': 10}, 'one': {'/content/dataset/1998.txt.utf-8': 778, '/content/dataset/pg1513.txt': 62, '/content/dataset/pg11.txt': 103, '/content/dataset/2160-0.txt': 356, '/content/dataset/345.txt.utf-8': 507, '/content/dataset/64317.txt.utf-8': 154, '/content/dataset/70817.txt.utf-8': 76, '/content/dataset/1080-0.txt': 17, '/content/dataset/pg2600.txt': 2134, '/content/dataset/70815.txt.utf-8': 245}, 'goals': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'way': {'/content/dataset/1998.txt.utf-8': 117, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 58, '/content/dataset/2160-0.txt': 150, '/content/dataset/345.txt.utf-8': 233, '/content/dataset/64317.txt.utf-8': 77, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 494, '/content/dataset/70815.txt.utf-8': 111}, 'creating': {'/content/dataset/1998.txt.utf-8': 45, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 4, '/content/dataset/70815.txt.utf-8': 4}, 'old': {'/content/dataset/1998.txt.utf-8': 174, '/content/dataset/pg1513.txt': 30, '/content/dataset/pg11.txt': 21, '/content/dataset/2160-0.txt': 135, '/content/dataset/345.txt.utf-8': 192, '/content/dataset/64317.txt.utf-8': 86, '/content/dataset/70817.txt.utf-8': 29, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 835, '/content/dataset/70815.txt.utf-8': 10}, 'young': {'/content/dataset/1998.txt.utf-8': 26, '/content/dataset/pg1513.txt': 24, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 83, '/content/dataset/345.txt.utf-8': 54, '/content/dataset/64317.txt.utf-8': 45, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 445, '/content/dataset/70815.txt.utf-8': 20}, 'child': {'/content/dataset/1998.txt.utf-8': 53, '/content/dataset/pg1513.txt': 16, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 17, '/content/dataset/345.txt.utf-8': 76, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 109, '/content/dataset/70815.txt.utf-8': 3}, 'second': {'/content/dataset/1998.txt.utf-8': 28, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 27, '/content/dataset/345.txt.utf-8': 20, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 122, '/content/dataset/70815.txt.utf-8': 9}, 'great': {'/content/dataset/1998.txt.utf-8': 242, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 39, '/content/dataset/2160-0.txt': 263, '/content/dataset/345.txt.utf-8': 183, '/content/dataset/64317.txt.utf-8': 31, '/content/dataset/70817.txt.utf-8': 16, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 265, '/content/dataset/70815.txt.utf-8': 165}, 'return': {'/content/dataset/1998.txt.utf-8': 38, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 51, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 100, '/content/dataset/70815.txt.utf-8': 48}, 'home': {'/content/dataset/1998.txt.utf-8': 46, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 35, '/content/dataset/345.txt.utf-8': 84, '/content/dataset/64317.txt.utf-8': 27, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 173, '/content/dataset/70815.txt.utf-8': 8}, 'things': {'/content/dataset/1998.txt.utf-8': 168, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 33, '/content/dataset/2160-0.txt': 26, '/content/dataset/345.txt.utf-8': 178, '/content/dataset/64317.txt.utf-8': 32, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 233,

16}, '/content/dataset/70815.txt.utf-8': 16}, '/content/dataset/1998.txt.utf-8': 16},
3, '/content/dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 33, '/content/dataset/
2160-0.txt': 100, '/content/dataset/345.txt.utf-8': 140, '/content/dataset/64317.txt.utf
-8': 45, '/content/dataset/70817.txt.utf-8': 24, '/content/dataset/1080-0.txt': 1, '/con
tent/dataset/pg2600.txt': 346, '/content/dataset/70815.txt.utf-8': 43}, 'talk': {'/con
tent/dataset/1998.txt.utf-8': 34, '/content/dataset/pg1513.txt': 15, '/content/dataset/pg1
1.txt': 14, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 36, '/c
ontent/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 3, '/content/d
ataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 254, '/content/dataset/70815.txt.ut
f-8': 5}, 'out': {'/content/dataset/1998.txt.utf-8': 254, '/content/dataset/pg1513.txt':
54, '/content/dataset/pg11.txt': 118, '/content/dataset/2160-0.txt': 226, '/content/dat
aset/345.txt.utf-8': 431, '/content/dataset/64317.txt.utf-8': 214, '/content/dataset/708
17.txt.utf-8': 107, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 124
0, '/content/dataset/70815.txt.utf-8': 117}, 'service': {'/content/dataset/1998.txt.utf-
8': 14, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/data
set/2160-0.txt': 62, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.u
tf-8': 4, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 2, '/con
tent/dataset/pg2600.txt': 130, '/content/dataset/70815.txt.utf-8': 2}, 'man': {'/con
tent/dataset/1998.txt.utf-8': 414, '/content/dataset/pg1513.txt': 71, '/content/dataset/pg1
1.txt': 5, '/content/dataset/2160-0.txt': 184, '/content/dataset/345.txt.utf-8': 256, '/
content/dataset/64317.txt.utf-8': 114, '/content/dataset/70817.txt.utf-8': 35, '/con
tent/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 1189, '/content/dataset/70815.t
xt.utf-8': 280}, 'among': {'/content/dataset/1998.txt.utf-8': 69, '/content/dataset/pg15
13.txt': 6, '/content/dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 93, '/con
tent/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/7
0817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 21
6, '/content/dataset/70815.txt.utf-8': 57}, 'how': {'/content/dataset/1998.txt.utf-8': 2
68, '/content/dataset/pg1513.txt': 62, '/content/dataset/pg11.txt': 71, '/content/datase
t/2160-0.txt': 107, '/content/dataset/345.txt.utf-8': 225, '/content/dataset/64317.txt.u
tf-8': 48, '/content/dataset/70817.txt.utf-8': 20, '/content/dataset/1080-0.txt': 8, '/c
ontent/dataset/pg2600.txt': 1010, '/content/dataset/70815.txt.utf-8': 72}, 'came': {'/co
ntent/dataset/1998.txt.utf-8': 87, '/content/dataset/pg1513.txt': 13, '/content/dataset/
pg11.txt': 40, '/content/dataset/2160-0.txt': 72, '/content/dataset/345.txt.utf-8': 309,
'/content/dataset/64317.txt.utf-8': 108, '/content/dataset/70817.txt.utf-8': 30, '/cont
ent/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 683, '/content/dataset/70815.
txt.utf-8': 92}, 'into': {'/content/dataset/1998.txt.utf-8': 257, '/content/dataset/pg15
13.txt': 8, '/content/dataset/pg11.txt': 67, '/content/dataset/2160-0.txt': 299, '/con
tent/dataset/345.txt.utf-8': 288, '/content/dataset/64317.txt.utf-8': 168, '/content/datas
et/70817.txt.utf-8': 80, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.tx
t': 1004, '/content/dataset/70815.txt.utf-8': 182}, 'being': {'/content/dataset/1998.tx
t.utf-8': 50, '/content/dataset/pg1513.txt': 26, '/content/dataset/pg11.txt': 19, '/cont
ent/dataset/2160-0.txt': 225, '/content/dataset/345.txt.utf-8': 84, '/content/dataset/64
317.txt.utf-8': 16, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.tx
t': 17, '/content/dataset/pg2600.txt': 463, '/content/dataset/70815.txt.utf-8': 60}, 'm
y': {'/content/dataset/1998.txt.utf-8': 981, '/content/dataset/pg1513.txt': 356, '/con
tent/dataset/pg11.txt': 58, '/content/dataset/2160-0.txt': 1322, '/content/dataset/345.tx
t.utf-8': 1262, '/content/dataset/64317.txt.utf-8': 244, '/content/dataset/70817.txt.utf
-8': 75, '/content/dataset/1080-0.txt': 20, '/content/dataset/pg2600.txt': 1226, '/con
tent/dataset/70815.txt.utf-8': 94}, 'work': {'/content/dataset/1998.txt.utf-8': 83, '/con
tent/dataset/pg1513.txt': 49, '/content/dataset/pg11.txt': 53, '/content/dataset/2160-0.t
xt': 67, '/content/dataset/345.txt.utf-8': 191, '/content/dataset/64317.txt.utf-8': 56,
'/content/dataset/70817.txt.utf-8': 54, '/content/dataset/1080-0.txt': 51, '/content/d
ataset/pg2600.txt': 166, '/content/dataset/70815.txt.utf-8': 56}, 'his': {'/content/datas
et/1998.txt.utf-8': 898, '/content/dataset/pg1513.txt': 134, '/content/dataset/pg11.tx
t': 95, '/content/dataset/2160-0.txt': 2023, '/content/dataset/345.txt.utf-8': 1471, '/c
ontent/dataset/64317.txt.utf-8': 494, '/content/dataset/70817.txt.utf-8': 282, '/con
tent/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 7984, '/content/dataset/70815.
txt.utf-8': 926}, 'individual': {'/content/dataset/1998.txt.utf-8': 12, '/content/datase
t/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 25, '/c
ontent/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 5, '/content/datas
et/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt':
18, '/content/dataset/70815.txt.utf-8': 7}, 'above': {'/content/dataset/1998.txt.utf-
8': 58, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 3, '/content/d
aset/2160-0.txt': 80, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.
utf-8': 7, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 4, '/co
ntent/dataset/pg2600.txt': 146, '/content/dataset/70815.txt.utf-8': 99}, 'however': {'/c

t/dataset/1998.txt.utf-8': 373, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 21, '/content/dataset/2160-0.txt': 100, '/content/dataset/345.txt.utf-8': 103, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 19, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 101, '/content/dataset/70815.txt.utf-8': 23}, 'there': {'/content/dataset/1998.txt.utf-8': 503, '/content/dataset/pg1513.txt': 70, '/content/dataset/pg11.txt': 100, '/content/dataset/2160-0.txt': 316, '/content/dataset/345.txt.utf-8': 782, '/content/dataset/64317.txt.utf-8': 221, '/content/dataset/70817.txt.utf-8': 85, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 1562, '/content/dataset/70815.txt.utf-8': 299}, 'figure': {'/content/dataset/1998.txt.utf-8': 9, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 24, '/content/dataset/345.txt.utf-8': 22, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 70, '/content/dataset/70815.txt.utf-8': 6}, 'from': {'/content/dataset/1998.txt.utf-8': 454, '/content/dataset/pg1513.txt': 101, '/content/dataset/pg11.txt': 52, '/content/dataset/2160-0.txt': 666, '/content/dataset/345.txt.utf-8': 640, '/content/dataset/64317.txt.utf-8': 247, '/content/dataset/70817.txt.utf-8': 92, '/content/dataset/1080-0.txt': 29, '/content/dataset/pg2600.txt': 2709, '/content/dataset/70815.txt.utf-8': 501}, 'very': {'/content/dataset/1998.txt.utf-8': 79, '/content/dataset/pg1513.txt': 26, '/content/dataset/pg11.txt': 131, '/content/dataset/2160-0.txt': 360, '/content/dataset/345.txt.utf-8': 280, '/content/dataset/64317.txt.utf-8': 61, '/content/dataset/70817.txt.utf-8': 32, '/content/dataset/1080-0.txt': 11, '/content/dataset/pg2600.txt': 804, '/content/dataset/70815.txt.utf-8': 74}, 'he': {'/content/dataset/1998.txt.utf-8': 1101, '/content/dataset/pg1513.txt': 136, '/content/dataset/pg11.txt': 125, '/content/dataset/2160-0.txt': 2168, '/content/dataset/345.txt.utf-8': 2581, '/content/dataset/64317.txt.utf-8': 857, '/content/dataset/70817.txt.utf-8': 557, '/content/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 10007, '/content/dataset/70815.txt.utf-8': 1477}, 'told': {'/content/dataset/1998.txt.utf-8': 13, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 79, '/content/dataset/345.txt.utf-8': 144, '/content/dataset/64317.txt.utf-8': 45, '/content/dataset/70817.txt.utf-8': 12, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 409, '/content/dataset/70815.txt.utf-8': 27}, 'me': {'/content/dataset/1998.txt.utf-8': 829, '/content/dataset/pg1513.txt': 266, '/content/dataset/pg11.txt': 64, '/content/dataset/2160-0.txt': 520, '/content/dataset/345.txt.utf-8': 1455, '/content/dataset/64317.txt.utf-8': 327, '/content/dataset/70817.txt.utf-8': 61, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 1275, '/content/dataset/70815.txt.utf-8': 105}, 'that': {'/content/dataset/1998.txt.utf-8': 1364, '/content/dataset/pg1513.txt': 369, '/content/dataset/pg11.txt': 315, '/content/dataset/2160-0.txt': 1901, '/content/dataset/345.txt.utf-8': 2502, '/content/dataset/64317.txt.utf-8': 658, '/content/dataset/70817.txt.utf-8': 273, '/content/dataset/1080-0.txt': 54, '/content/dataset/pg2600.txt': 8205, '/content/dataset/70815.txt.utf-8': 1447}, 'even': {'/content/dataset/1998.txt.utf-8': 255, '/content/dataset/pg1513.txt': 14, '/content/dataset/pg11.txt': 20, '/content/dataset/2160-0.txt': 181, '/content/dataset/345.txt.utf-8': 244, '/content/dataset/64317.txt.utf-8': 49, '/content/dataset/70817.txt.utf-8': 41, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 512, '/content/dataset/70815.txt.utf-8': 98}, 'as': {'/content/dataset/1998.txt.utf-8': 548, '/content/dataset/pg1513.txt': 168, '/content/dataset/pg11.txt': 274, '/content/dataset/2160-0.txt': 1174, '/content/dataset/345.txt.utf-8': 1602, '/content/dataset/64317.txt.utf-8': 281, '/content/dataset/70817.txt.utf-8': 133, '/content/dataset/1080-0.txt': 50, '/content/dataset/pg2600.txt': 4035, '/content/dataset/70815.txt.utf-8': 690}, 'him': {'/content/dataset/1998.txt.utf-8': 406, '/content/dataset/pg1513.txt': 98, '/content/dataset/pg11.txt': 42, '/content/dataset/2160-0.txt': 684, '/content/dataset/345.txt.utf-8': 953, '/content/dataset/64317.txt.utf-8': 278, '/content/dataset/70817.txt.utf-8': 108, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 4637, '/content/dataset/70815.txt.utf-8': 502}, 'different': {'/content/dataset/1998.txt.utf-8': 19, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 69, '/content/dataset/345.txt.utf-8': 23, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 146, '/content/dataset/70815.txt.utf-8': 8}, 'life': {'/content/dataset/1998.txt.utf-8': 197, '/content/dataset/pg1513.txt': 25, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 138, '/content/dataset/345.txt.utf-8': 150, '/content/dataset/64317.txt.utf-8': 31, '/content/dataset/70817.txt.utf-8': 21, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 618, '/content/dataset/70815.txt.utf-8': 64}, 'would': {'/content/dataset/1998.txt.utf-8': 261, '/content/dataset/pg1513.txt': 55, '/content/dataset/pg11.txt': 78, '/content/dataset/2160-0.txt': 446, '/content/dataset/345.txt.utf-8': 432, '/content/dataset/64317.txt.utf-8': 53, '/content/dataset/70817.txt.utf-8': 88, '/content/dataset/1080-0.txt': 18, '/content/dataset/pg2600.txt': 1366, '/content/dataset/70815.txt.utf-8': 219}, 'but': {'/content/dataset/1998.txt.utf-8': 197, '/content/dataset/pg1513.txt': 25, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 138, '/content/dataset/345.txt.utf-8': 150, '/content/dataset/64317.txt.utf-8': 31, '/content/dataset/70817.txt.utf-8': 21, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 618, '/content/dataset/70815.txt.utf-8': 64}

t/1998.txt.utf-8': 715, '/content/dataset/pg1513.txt': 185, '/content/dataset/pg11.txt': 174, '/content/dataset/2160-0.txt': 1102, '/content/dataset/345.txt.utf-8': 1075, '/content/dataset/64317.txt.utf-8': 260, '/content/dataset/70817.txt.utf-8': 153, '/content/dataset/1080-0.txt': 13, '/content/dataset/pg2600.txt': 4054, '/content/dataset/70815.txt.utf-8': 559}, 'end': {'/content/dataset/1998.txt.utf-8': 45, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 19, '/content/dataset/2160-0.txt': 39, '/content/dataset/345.txt.utf-8': 77, '/content/dataset/64317.txt.utf-8': 22, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 185, '/content/dataset/70815.txt.utf-8': 38}, 'subject': {'/content/dataset/1998.txt.utf-8': 9, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 7, '/content/dataset/2160-0.txt': 57, '/content/dataset/345.txt.utf-8': 24, '/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 87, '/content/dataset/70815.txt.utf-8': 6}, 'do': {'/content/dataset/1998.txt.utf-8': 527, '/content/dataset/pg1513.txt': 105, '/content/dataset/pg11.txt': 97, '/content/dataset/2160-0.txt': 113, '/content/dataset/345.txt.utf-8': 491, '/content/dataset/64317.txt.utf-8': 104, '/content/dataset/70817.txt.utf-8': 26, '/content/dataset/1080-0.txt': 19, '/content/dataset/pg2600.txt': 1058, '/content/dataset/70815.txt.utf-8': 165}, 'were': {'/content/dataset/1998.txt.utf-8': 104, '/content/dataset/pg1513.txt': 37, '/content/dataset/pg11.txt': 85, '/content/dataset/2160-0.txt': 339, '/content/dataset/345.txt.utf-8': 552, '/content/dataset/64317.txt.utf-8': 173, '/content/dataset/70817.txt.utf-8': 108, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 2425, '/content/dataset/70815.txt.utf-8': 482}, 'take': {'/content/dataset/1998.txt.utf-8': 67, '/content/dataset/pg1513.txt': 40, '/content/dataset/pg11.txt': 22, '/content/dataset/2160-0.txt': 144, '/content/dataset/345.txt.utf-8': 150, '/content/dataset/64317.txt.utf-8': 28, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 352, '/content/dataset/70815.txt.utf-8': 61}, 'every': {'/content/dataset/1998.txt.utf-8': 127, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 215, '/content/dataset/345.txt.utf-8': 84, '/content/dataset/64317.txt.utf-8': 28, '/content/dataset/70817.txt.utf-8': 21, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 334, '/content/dataset/70815.txt.utf-8': 52}, 'them': {'/content/dataset/1998.txt.utf-8': 336, '/content/dataset/pg1513.txt': 43, '/content/dataset/pg11.txt': 88, '/content/dataset/2160-0.txt': 210, '/content/dataset/345.txt.utf-8': 474, '/content/dataset/64317.txt.utf-8': 82, '/content/dataset/70817.txt.utf-8': 36, '/content/dataset/1080-0.txt': 16, '/content/dataset/pg2600.txt': 1529, '/content/dataset/70815.txt.utf-8': 398}, 'was': {'/content/dataset/1998.txt.utf-8': 362, '/content/dataset/pg1513.txt': 51, '/content/dataset/pg11.txt': 354, '/content/dataset/2160-0.txt': 1401, '/content/dataset/345.txt.utf-8': 1881, '/content/dataset/64317.txt.utf-8': 770, '/content/dataset/70817.txt.utf-8': 441, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 7365, '/content/dataset/70815.txt.utf-8': 1094}, 'years': {'/content/dataset/1998.txt.utf-8': 27, '/content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 84, '/content/dataset/345.txt.utf-8': 30, '/content/dataset/64317.txt.utf-8': 40, '/content/dataset/70817.txt.utf-8': 16, '/content/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 119, '/content/dataset/70815.txt.utf-8': 11}, 'also': {'/content/dataset/1998.txt.utf-8': 346, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 42, '/content/dataset/345.txt.utf-8': 44, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 270, '/content/dataset/70815.txt.utf-8': 20}, 'care': {'/content/dataset/1998.txt.utf-8': 10, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 47, '/content/dataset/345.txt.utf-8': 52, '/content/dataset/64317.txt.utf-8': 14, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 48, '/content/dataset/70815.txt.utf-8': 10}, 'meet': {'/content/dataset/1998.txt.utf-8': 20, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 21, '/content/dataset/345.txt.utf-8': 42, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 107, '/content/dataset/70815.txt.utf-8': 17}, 'put': {'/content/dataset/1998.txt.utf-8': 60, '/content/dataset/pg1513.txt': 17, '/content/dataset/pg11.txt': 31, '/content/dataset/2160-0.txt': 86, '/content/dataset/345.txt.utf-8': 107, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 271, '/content/dataset/70815.txt.utf-8': 10}, 'forth': {'/content/dataset/1998.txt.utf-8': 39, '/content/dataset/pg1513.txt': 20, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 34, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 11, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 19, '/content/dataset/70815.txt.utf-8': 16}, 'clearly': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}

nt/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt': 2, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 92, '/content/dataset/70815.txt.utf-8': 2}, 'we': {'/content/dataset/1998.txt.utf-8': 294, '/content/dataset/pg1513.txt': 85, '/content/dataset/pg11.txt': 42, '/content/dataset/2160-0.txt': 562, '/content/dataset/345.txt.utf-8': 1561, '/content/dataset/64317.txt.utf-8': 243, '/content/dataset/70817.txt.utf-8': 96, '/content/dataset/1080-0.txt': 18, '/content/dataset/pg2600.txt': 1078, '/content/dataset/70815.txt.utf-8': 363}, 'following': {'/content/dataset/1998.txt.utf-8': 11, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 21, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 75, '/content/dataset/70815.txt.utf-8': 25}, 'occure': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 24, '/content/dataset/70815.txt.utf-8': 1}, 'can': {'/content/dataset/1998.txt.utf-8': 115, '/content/dataset/pg1513.txt': 49, '/content/dataset/pg11.txt': 69, '/content/dataset/2160-0.txt': 169, '/content/dataset/345.txt.utf-8': 418, '/content/dataset/64317.txt.utf-8': 75, '/content/dataset/70817.txt.utf-8': 22, '/content/dataset/1080-0.txt': 19, '/content/dataset/pg2600.txt': 674, '/content/dataset/70815.txt.utf-8': 79}, 'whole': {'/content/dataset/1998.txt.utf-8': 30, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 102, '/content/dataset/345.txt.utf-8': 73, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 532, '/content/dataset/70815.txt.utf-8': 15}, 'important': {'/content/dataset/1998.txt.utf-8': 19, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 13, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 131, '/content/dataset/70815.txt.utf-8': 5}, 'because': {'/content/dataset/1998.txt.utf-8': 100, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 16, '/content/dataset/2160-0.txt': 43, '/content/dataset/345.txt.utf-8': 56, '/content/dataset/64317.txt.utf-8': 46, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 491, '/content/dataset/70815.txt.utf-8': 43}, 'they': {'/content/dataset/1998.txt.utf-8': 665, '/content/dataset/pg1513.txt': 56, '/content/dataset/pg11.txt': 153, '/content/dataset/2160-0.txt': 502, '/content/dataset/345.txt.utf-8': 467, '/content/dataset/64317.txt.utf-8': 170, '/content/dataset/70817.txt.utf-8': 76, '/content/dataset/1080-0.txt': 30, '/content/dataset/pg2600.txt': 2255, '/content/dataset/70815.txt.utf-8': 678}, 'such': {'/content/dataset/1998.txt.utf-8': 183, '/content/dataset/pg1513.txt': 52, '/content/dataset/pg11.txt': 47, '/content/dataset/2160-0.txt': 402, '/content/dataset/345.txt.utf-8': 233, '/content/dataset/64317.txt.utf-8': 27, '/content/dataset/70817.txt.utf-8': 16, '/content/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 597, '/content/dataset/70815.txt.utf-8': 94}, 'number': {'/content/dataset/1998.txt.utf-8': 12, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 58, '/content/dataset/345.txt.utf-8': 21, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 14, '/content/dataset/pg2600.txt': 79, '/content/dataset/70815.txt.utf-8': 17}, 'which': {'/content/dataset/1998.txt.utf-8': 470, '/content/dataset/pg1513.txt': 70, '/content/dataset/pg11.txt': 56, '/content/dataset/2160-0.txt': 1024, '/content/dataset/345.txt.utf-8': 670, '/content/dataset/64317.txt.utf-8': 72, '/content/dataset/70817.txt.utf-8': 132, '/content/dataset/1080-0.txt': 30, '/content/dataset/pg2600.txt': 2064, '/content/dataset/70815.txt.utf-8': 400}, 'be': {'/content/dataset/1998.txt.utf-8': 788, '/content/dataset/pg1513.txt': 227, '/content/dataset/pg11.txt': 167, '/content/dataset/2160-0.txt': 987, '/content/dataset/345.txt.utf-8': 1133, '/content/dataset/64317.txt.utf-8': 141, '/content/dataset/70817.txt.utf-8': 104, '/content/dataset/1080-0.txt': 70, '/content/dataset/pg2600.txt': 2455, '/content/dataset/70815.txt.utf-8': 325}, 'am': {'/content/dataset/1998.txt.utf-8': 189, '/content/dataset/pg1513.txt': 60, '/content/dataset/pg11.txt': 16, '/content/dataset/2160-0.txt': 221, '/content/dataset/345.txt.utf-8': 337, '/content/dataset/64317.txt.utf-8': 16, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 551, '/content/dataset/70815.txt.utf-8': 87}, 'only': {'/content/dataset/1998.txt.utf-8': 247, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 52, '/content/dataset/2160-0.txt': 115, '/content/dataset/345.txt.utf-8': 289, '/content/dataset/64317.txt.utf-8': 55, '/content/dataset/70817.txt.utf-8': 34, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 1300, '/content/dataset/70815.txt.utf-8': 128}, 'people': {'/content/dataset/1998.txt.utf-8': 141, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 16, '/content/dataset/2160-0.txt': 1

13, '/content/dataset/345.txt.utf-8': 47, '/content/dataset/64317.txt.utf-8': 66, '/content/dataset/70817.txt.utf-8': 12, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 560, '/content/dataset/70815.txt.utf-8': 61}, 'any': {'/content/dataset/1998.txt.utf-8': 113, '/content/dataset/pg1513.txt': 53, '/content/dataset/pg11.txt': 76, '/content/dataset/2160-0.txt': 338, '/content/dataset/345.txt.utf-8': 322, '/content/dataset/64317.txt.utf-8': 105, '/content/dataset/70817.txt.utf-8': 68, '/content/dataset/1080-0.txt': 45, '/content/dataset/pg2600.txt': 503, '/content/dataset/70815.txt.utf-8': 162}, 'means': {'/content/dataset/1998.txt.utf-8': 32, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 63, '/content/dataset/345.txt.utf-8': 27, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 73, '/content/dataset/70815.txt.utf-8': 35}, 'their': {'/content/dataset/1998.txt.utf-8': 427, '/content/dataset/pg1513.txt': 47, '/content/dataset/pg11.txt': 51, '/content/dataset/2160-0.txt': 479, '/content/dataset/345.txt.utf-8': 219, '/content/dataset/64317.txt.utf-8': 56, '/content/dataset/70817.txt.utf-8': 37, '/content/dataset/1080-0.txt': 38, '/content/dataset/pg2600.txt': 1440, '/content/dataset/70815.txt.utf-8': 376}, 'help': {'/content/dataset/1998.txt.utf-8': 27, '/content/dataset/pg1513.txt': 21, '/content/dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 55, '/content/dataset/345.txt.utf-8': 104, '/content/dataset/64317.txt.utf-8': 10, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 144, '/content/dataset/70815.txt.utf-8': 9}, 'might': {'/content/dataset/1998.txt.utf-8': 44, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 28, '/content/dataset/2160-0.txt': 137, '/content/dataset/345.txt.utf-8': 158, '/content/dataset/64317.txt.utf-8': 26, '/content/dataset/70817.txt.utf-8': 22, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 315, '/content/dataset/70815.txt.utf-8': 163}, 'who': {'/content/dataset/1998.txt.utf-8': 541, '/content/dataset/pg1513.txt': 38, '/content/dataset/pg11.txt': 67, '/content/dataset/2160-0.txt': 618, '/content/dataset/345.txt.utf-8': 289, '/content/dataset/64317.txt.utf-8': 117, '/content/dataset/70817.txt.utf-8': 20, '/content/dataset/1080-0.txt': 27, '/content/dataset/pg2600.txt': 2170, '/content/dataset/70815.txt.utf-8': 166}, 'both': {'/content/dataset/1998.txt.utf-8': 30, '/content/dataset/pg1513.txt': 24, '/content/dataset/pg11.txt': 15, '/content/dataset/2160-0.txt': 72, '/content/dataset/345.txt.utf-8': 87, '/content/dataset/64317.txt.utf-8': 14, '/content/dataset/70817.txt.utf-8': 12, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 262, '/content/dataset/70815.txt.utf-8': 29}, 'than': {'/content/dataset/1998.txt.utf-8': 206, '/content/dataset/pg1513.txt': 50, '/content/dataset/pg11.txt': 27, '/content/dataset/2160-0.txt': 280, '/content/dataset/345.txt.utf-8': 223, '/content/dataset/64317.txt.utf-8': 37, '/content/dataset/70817.txt.utf-8': 37, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 536, '/content/dataset/70815.txt.utf-8': 151}, 'those': {'/content/dataset/1998.txt.utf-8': 206, '/content/dataset/pg1513.txt': 16, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 178, '/content/dataset/345.txt.utf-8': 83, '/content/dataset/64317.txt.utf-8': 28, '/content/dataset/70817.txt.utf-8': 20, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 662, '/content/dataset/70815.txt.utf-8': 56}, 'owed': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 4, '/content/dataset/70815.txt.utf-8': 2}, 'men': {'/content/dataset/1998.txt.utf-8': 230, '/content/dataset/pg1513.txt': 30, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 40, '/content/dataset/345.txt.utf-8': 151, '/content/dataset/64317.txt.utf-8': 43, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 792, '/content/dataset/70815.txt.utf-8': 167}, 'form': {'/content/dataset/1998.txt.utf-8': 23, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 19, '/content/dataset/345.txt.utf-8': 24, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 59, '/content/dataset/70815.txt.utf-8': 13}, 'an': {'/content/dataset/1998.txt.utf-8': 245, '/content/dataset/pg1513.txt': 76, '/content/dataset/pg11.txt': 62, '/content/dataset/2160-0.txt': 565, '/content/dataset/345.txt.utf-8': 365, '/content/dataset/64317.txt.utf-8': 193, '/content/dataset/70817.txt.utf-8': 105, '/content/dataset/1080-0.txt': 14, '/content/dataset/pg2600.txt': 1636, '/content/dataset/70815.txt.utf-8': 250}, 'already': {'/content/dataset/1998.txt.utf-8': 79, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 39, '/content/dataset/345.txt.utf-8': 65, '/content/dataset/64317.txt.utf-8': 16, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 326, '/content/dataset/70815.txt.utf-8': 27}, 'should': {'/content/dataset/1998.txt.utf-8': 137, '/content/dataset/pg1513.txt': 51, '/content/dataset/pg11.txt': 29, '/content/dataset/2160-0.txt': 260, '/content/dataset/345.txt.utf-8': 199, '/content/dataset/64317.

txt.utf-8': 11, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 523, '/content/dataset/70815.txt.utf-8': 59}, 'its': {'/content/dataset/1998.txt.utf-8': 264, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 60, '/content/dataset/2160-0.txt': 97, '/content/dataset/345.txt.utf-8': 138, '/content/dataset/64317.txt.utf-8': 43, '/content/dataset/70817.txt.utf-8': 63, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 571, '/content/dataset/70815.txt.utf-8': 234}, 'produce': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 25, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 18, '/content/dataset/70815.txt.utf-8': 2}, 'nothing': {'/content/dataset/1998.txt.utf-8': 64, '/content/dataset/pg1513.txt': 14, '/content/dataset/pg11.txt': 34, '/content/dataset/2160-0.txt': 139, '/content/dataset/345.txt.utf-8': 112, '/content/dataset/64317.txt.utf-8': 35, '/content/dataset/70817.txt.utf-8': 22, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 473, '/content/dataset/70815.txt.utf-8': 42}, 'days': {'/content/dataset/1998.txt.utf-8': 26, '/content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 58, '/content/dataset/345.txt.utf-8': 61, '/content/dataset/64317.txt.utf-8': 23, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 198, '/content/dataset/70815.txt.utf-8': 8}, 'longer': {'/content/dataset/1998.txt.utf-8': 90, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 34, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 14, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 176, '/content/dataset/70815.txt.utf-8': 18}, 'types': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 2}, 'future': {'/content/dataset/1998.txt.utf-8': 49, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 11, '/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 77, '/content/dataset/70815.txt.utf-8': 155}, 'what': {'/content/dataset/1998.txt.utf-8': 444, '/content/dataset/pg1513.txt': 174, '/content/dataset/pg11.txt': 135, '/content/dataset/2160-0.txt': 289, '/content/dataset/345.txt.utf-8': 467, '/content/dataset/64317.txt.utf-8': 147, '/content/dataset/70817.txt.utf-8': 38, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 2399, '/content/dataset/70815.txt.utf-8': 176}, 'after': {'/content/dataset/1998.txt.utf-8': 80, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 43, '/content/dataset/2160-0.txt': 124, '/content/dataset/345.txt.utf-8': 210, '/content/dataset/64317.txt.utf-8': 89, '/content/dataset/70817.txt.utf-8': 50, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 841, '/content/dataset/70815.txt.utf-8': 79}, 'having': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 144, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 11, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 494, '/content/dataset/70815.txt.utf-8': 26}, 'our': {'/content/dataset/1998.txt.utf-8': 116, '/content/dataset/pg1513.txt': 64, '/content/dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 424, '/content/dataset/345.txt.utf-8': 411, '/content/dataset/64317.txt.utf-8': 28, '/content/dataset/70817.txt.utf-8': 19, '/content/dataset/1080-0.txt': 26, '/content/dataset/pg2600.txt': 626, '/content/dataset/70815.txt.utf-8': 80}, 'never': {'/content/dataset/1998.txt.utf-8': 64, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 46, '/content/dataset/2160-0.txt': 141, '/content/dataset/345.txt.utf-8': 128, '/content/dataset/64317.txt.utf-8': 64, '/content/dataset/70817.txt.utf-8': 29, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 409, '/content/dataset/70815.txt.utf-8': 94}, 'yet': {'/content/dataset/1998.txt.utf-8': 131, '/content/dataset/pg1513.txt': 47, '/content/dataset/pg11.txt': 24, '/content/dataset/2160-0.txt': 85, '/content/dataset/345.txt.utf-8': 156, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 21, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 347, '/content/dataset/70815.txt.utf-8': 75}, 'too': {'/content/dataset/1998.txt.utf-8': 189, '/content/dataset/pg1513.txt': 56, '/content/dataset/pg11.txt': 26, '/content/dataset/2160-0.txt': 70, '/content/dataset/345.txt.utf-8': 215, '/content/dataset/64317.txt.utf-8': 63, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 393, '/content/dataset/70815.txt.utf-8': 42}, 'each': {'/content/dataset/1998.txt.utf-8': 42, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 30, '/content/dataset/345.txt.utf-8': 87, '/content/dataset/64317.txt.utf-8': 34, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 1}

1, '/content/dataset/pg2600.txt': 183, '/content/dataset/70817.txt.utf-8': 47}, 'found': {'/content/dataset/1998.txt.utf-8': 83, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 35, '/content/dataset/2160-0.txt': 121, '/content/dataset/345.txt.utf-8': 155, '/content/dataset/64317.txt.utf-8': 38, '/content/dataset/70817.txt.utf-8': 25, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 222, '/content/dataset/70815.txt.utf-8': 76}, 'phrase': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 10, '/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 15, '/content/dataset/70815.txt.utf-8': 4}, 'has': {'/content/dataset/1998.txt.utf-8': 47, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 401, '/content/dataset/345.txt.utf-8': 347, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 709, '/content/dataset/70815.txt.utf-8': 68}, 'word': {'/content/dataset/1998.txt.utf-8': 57, '/content/dataset/pg1513.txt': 29, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 43, '/content/dataset/345.txt.utf-8': 66, '/content/dataset/64317.txt.utf-8': 24, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 235, '/content/dataset/70815.txt.utf-8': 15}, 'now': {'/content/dataset/1998.txt.utf-8': 313, '/content/dataset/pg1513.txt': 81, '/content/dataset/pg11.txt': 60, '/content/dataset/2160-0.txt': 199, '/content/dataset/345.txt.utf-8': 407, '/content/dataset/64317.txt.utf-8': 100, '/content/dataset/70817.txt.utf-8': 46, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 1332, '/content/dataset/70815.txt.utf-8': 147}, 'general': {'/content/dataset/1998.txt.utf-8': 9, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 51, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 506, '/content/dataset/70815.txt.utf-8': 9}, 'order': {'/content/dataset/1998.txt.utf-8': 40, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 65, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 295, '/content/dataset/70815.txt.utf-8': 5}, 'strength': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 46, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 119, '/content/dataset/70815.txt.utf-8': 16}, 'table': {'/content/dataset/1998.txt.utf-8': 20, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 18, '/content/dataset/2160-0.txt': 39, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 33, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 225, '/content/dataset/70815.txt.utf-8': 1}, 'must': {'/content/dataset/1998.txt.utf-8': 202, '/content/dataset/pg1513.txt': 50, '/content/dataset/pg11.txt': 51, '/content/dataset/2160-0.txt': 183, '/content/dataset/345.txt.utf-8': 450, '/content/dataset/64317.txt.utf-8': 42, '/content/dataset/70817.txt.utf-8': 25, '/content/dataset/1080-0.txt': 11, '/content/dataset/pg2600.txt': 482, '/content/dataset/70815.txt.utf-8': 110}, 'us': {'/content/dataset/1998.txt.utf-8': 154, '/content/dataset/pg1513.txt': 31, '/content/dataset/pg11.txt': 15, '/content/dataset/2160-0.txt': 155, '/content/dataset/345.txt.utf-8': 465, '/content/dataset/64317.txt.utf-8': 64, '/content/dataset/70817.txt.utf-8': 28, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 424, '/content/dataset/70815.txt.utf-8': 116}, 'hope': {'/content/dataset/1998.txt.utf-8': 39, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 68, '/content/dataset/345.txt.utf-8': 67, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 91, '/content/dataset/70815.txt.utf-8': 42}, 'just': {'/content/dataset/1998.txt.utf-8': 105, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 52, '/content/dataset/2160-0.txt': 42, '/content/dataset/345.txt.utf-8': 169, '/content/dataset/64317.txt.utf-8': 103, '/content/dataset/70817.txt.utf-8': 28, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 568, '/content/dataset/70815.txt.utf-8': 59}, 'so': {'/content/dataset/1998.txt.utf-8': 361, '/content/dataset/pg1513.txt': 149, '/content/dataset/pg11.txt': 153, '/content/dataset/2160-0.txt': 614, '/content/dataset/345.txt.utf-8': 1109, '/content/dataset/64317.txt.utf-8': 140, '/content/dataset/70817.txt.utf-8': 54, '/content/dataset/1080-0.txt': 15, '/content/dataset/pg2600.txt': 1902, '/content/dataset/70815.txt.utf-8': 233}, 'good': {'/content/dataset/1998.txt.utf-8': 294, '/content/dataset/pg1513.txt': 85, '/content/dataset/pg11.txt': 27, '/content/dataset/2160-0.txt': 274, '/content/dataset/345.txt.utf-8': 258, '/content/dataset/64317.txt.utf-8': 30, '/content/dataset/70817.txt.utf-8': 15, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 497, '/content/dataset/70815.txt.utf-

'txt': 46}, 'some': {'/content/dataset/1998.txt.utf-8': 73, '/content/dataset/pg1513.txt.utf-8': 58, '/content/dataset/pg11.txt': 51, '/content/dataset/2160-0.txt': 350, '/content/dataset/345.txt.utf-8': 444, '/content/dataset/64317.txt.utf-8': 104, '/content/dataset/70817.txt.utf-8': 76, '/content/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 847, '/content/dataset/70815.txt.utf-8': 103}, 'nor': {'/content/dataset/1998.txt.utf-8': 38, '/content/dataset/pg1513.txt': 20, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 90, '/content/dataset/345.txt.utf-8': 35, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 177, '/content/dataset/70815.txt.utf-8': 85}, 'know': {'/content/dataset/1998.txt.utf-8': 131, '/content/dataset/pg1513.txt': 39, '/content/dataset/pg11.txt': 88, '/content/dataset/2160-0.txt': 182, '/content/dataset/345.txt.utf-8': 392, '/content/dataset/64317.txt.utf-8': 100, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 847, '/content/dataset/70815.txt.utf-8': 101}, 'possibility': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 60, '/content/dataset/70815.txt.utf-8': 11}, 'present': {'/content/dataset/1998.txt.utf-8': 68, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 96, '/content/dataset/345.txt.utf-8': 57, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 125, '/content/dataset/70815.txt.utf-8': 7}, 'could': {'/content/dataset/1998.txt.utf-8': 118, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 75, '/content/dataset/2160-0.txt': 248, '/content/dataset/345.txt.utf-8': 492, '/content/dataset/64317.txt.utf-8': 62, '/content/dataset/70817.txt.utf-8': 64, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 1116, '/content/dataset/70815.txt.utf-8': 247}, 'physical': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 10, '/content/dataset/64317.txt.utf-8': 10, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 36, '/content/dataset/70815.txt.utf-8': 9}, 'provided': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 30, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 12, '/content/dataset/70815.txt.utf-8': 4}, 'through': {'/content/dataset/1998.txt.utf-8': 80, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 15, '/content/dataset/2160-0.txt': 75, '/content/dataset/345.txt.utf-8': 216, '/content/dataset/64317.txt.utf-8': 68, '/content/dataset/70817.txt.utf-8': 43, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 379, '/content/dataset/70815.txt.utf-8': 107}, 'well': {'/content/dataset/1998.txt.utf-8': 189, '/content/dataset/pg1513.txt': 62, '/content/dataset/pg11.txt': 63, '/content/dataset/2160-0.txt': 196, '/content/dataset/345.txt.utf-8': 245, '/content/dataset/64317.txt.utf-8': 70, '/content/dataset/70817.txt.utf-8': 24, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 746, '/content/dataset/70815.txt.utf-8': 82}, 'time': {'/content/dataset/1998.txt.utf-8': 159, '/content/dataset/pg1513.txt': 32, '/content/dataset/pg11.txt': 71, '/content/dataset/2160-0.txt': 217, '/content/dataset/345.txt.utf-8': 390, '/content/dataset/64317.txt.utf-8': 83, '/content/dataset/70817.txt.utf-8': 37, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 929, '/content/dataset/70815.txt.utf-8': 141}, 'made': {'/content/dataset/1998.txt.utf-8': 45, '/content/dataset/pg1513.txt': 24, '/content/dataset/pg11.txt': 30, '/content/dataset/2160-0.txt': 229, '/content/dataset/345.txt.utf-8': 196, '/content/dataset/64317.txt.utf-8': 63, '/content/dataset/70817.txt.utf-8': 35, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 461, '/content/dataset/70815.txt.utf-8': 75}, 'more': {'/content/dataset/1998.txt.utf-8': 243, '/content/dataset/pg1513.txt': 68, '/content/dataset/pg11.txt': 49, '/content/dataset/2160-0.txt': 286, '/content/dataset/345.txt.utf-8': 377, '/content/dataset/64317.txt.utf-8': 75, '/content/dataset/70817.txt.utf-8': 40, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 1058, '/content/dataset/70815.txt.utf-8': 121}, 'two': {'/content/dataset/1998.txt.utf-8': 48, '/content/dataset/pg1513.txt': 20, '/content/dataset/pg11.txt': 38, '/content/dataset/2160-0.txt': 151, '/content/dataset/345.txt.utf-8': 133, '/content/dataset/64317.txt.utf-8': 73, '/content/dataset/70817.txt.utf-8': 31, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 595, '/content/dataset/70815.txt.utf-8': 122}, 'effect': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 53, '/content/dataset/345.txt.utf-8': 28, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 65, '/content/dataset/70815.txt.utf-8': 19}, 'another': {'/content/dataset/1998.txt.utf-8': 76, '/'

content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 144, '/content/dataset/345.txt.utf-8': 90, '/content/dataset/64317.txt.utf-8': 36, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 620, '/content/dataset/70815.txt.utf-8': 79}, 'within': {'/content/dataset/1998.txt.utf-8': 11, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 54, '/content/dataset/345.txt.utf-8': 48, '/content/dataset/64317.txt.utf-8': 10, '/content/dataset/70817.txt.utf-8': 14, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 86, '/content/dataset/70815.txt.utf-8': 63}, 'shall': {'/content/dataset/1998.txt.utf-8': 204, '/content/dataset/pg1513.txt': 112, '/content/dataset/pg11.txt': 27, '/content/dataset/2160-0.txt': 223, '/content/dataset/345.txt.utf-8': 428, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 249, '/content/dataset/70815.txt.utf-8': 96}, 'appear': {'/content/dataset/1998.txt.utf-8': 11, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 28, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 53, '/content/dataset/70815.txt.utf-8': 4}, 'creation': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 5, '/content/dataset/70815.txt.utf-8': 4}, 'concerning': {'/content/dataset/1998.txt.utf-8': 34, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 10, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 30, '/content/dataset/70815.txt.utf-8': 6}, 'certain': {'/content/dataset/1998.txt.utf-8': 20, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 50, '/content/dataset/345.txt.utf-8': 41, '/content/dataset/64317.txt.utf-8': 15, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 139, '/content/dataset/70815.txt.utf-8': 21}, 'understand': {'/content/dataset/1998.txt.utf-8': 32, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 7, '/content/dataset/2160-0.txt': 38, '/content/dataset/345.txt.utf-8': 56, '/content/dataset/64317.txt.utf-8': 15, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 350, '/content/dataset/70815.txt.utf-8': 29}, 'depends': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 19, '/content/dataset/70815.txt.utf-8': 1}, 'express': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 62, '/content/dataset/70815.txt.utf-8': 2}, 'require': {'/content/dataset/1998.txt.utf-8': 9, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 6, '/content/dataset/70815.txt.utf-8': 4}, 'whose': {'/content/dataset/1998.txt.utf-8': 43, '/content/dataset/pg1513.txt': 15, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 69, '/content/dataset/345.txt.utf-8': 30, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 114, '/content/dataset/70815.txt.utf-8': 29}, 'non': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 3, '/content/dataset/70815.txt.utf-8': 1}, 'thing': {'/content/dataset/1998.txt.utf-8': 91, '/content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 49, '/content/dataset/2160-0.txt': 79, '/content/dataset/345.txt.utf-8': 97, '/content/dataset/64317.txt.utf-8': 36, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 229, '/content/dataset/70815.txt.utf-8': 32}, 'purpose': {'/content/dataset/1998.txt.utf-8': 27, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 37, '/content/dataset/345.txt.utf-8': 36, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 50, '/content/dataset/70815.txt.utf-8': 17}, 'again': {'/content/dataset/1998.txt.utf-8': 160, '/content/dataset/pg1513.txt': 27, '/content/dataset/pg11.txt': 83, '/content/dataset/2160-0.txt': 51, '/content/dataset/3

45.txt.utf-8': 246, '/content/dataset/64317.txt.utf-8': 57, '/content/dataset/70817.txt.utf-8': 35, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 709, '/content/dataset/70815.txt.utf-8': 83}, 'long': {'/content/dataset/1998.txt.utf-8': 169, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 34, '/content/dataset/2160-0.txt': 104, '/content/dataset/345.txt.utf-8': 149, '/content/dataset/64317.txt.utf-8': 63, '/content/dataset/70817.txt.utf-8': 29, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 545, '/content/dataset/70815.txt.utf-8': 108}, 'perhaps': {'/content/dataset/1998.txt.utf-8': 70, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 17, '/content/dataset/2160-0.txt': 59, '/content/dataset/345.txt.utf-8': 54, '/content/dataset/64317.txt.utf-8': 19, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 118, '/content/dataset/70815.txt.utf-8': 65}, 'brought': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 62, '/content/dataset/345.txt.utf-8': 36, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 211, '/content/dataset/70815.txt.utf-8': 21}, 'seem': {'/content/dataset/1998.txt.utf-8': 34, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 28, '/content/dataset/345.txt.utf-8': 65, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 83, '/content/dataset/70815.txt.utf-8': 14}, 'countries': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 4}, 'known': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 25, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 147, '/content/dataset/70815.txt.utf-8': 24}, 'possession': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 18, '/content/dataset/70815.txt.utf-8': 2}, 'day': {'/content/dataset/1998.txt.utf-8': 194, '/content/dataset/pg1513.txt': 50, '/content/dataset/pg11.txt': 29, '/content/dataset/2160-0.txt': 154, '/content/dataset/345.txt.utf-8': 224, '/content/dataset/64317.txt.utf-8': 45, '/content/dataset/70817.txt.utf-8': 23, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 567, '/content/dataset/70815.txt.utf-8': 14}, 'full': {'/content/dataset/1998.txt.utf-8': 48, '/content/dataset/pg1513.txt': 24, '/content/dataset/pg11.txt': 19, '/content/dataset/2160-0.txt': 50, '/content/dataset/345.txt.utf-8': 102, '/content/dataset/64317.txt.utf-8': 34, '/content/dataset/70817.txt.utf-8': 22, '/content/dataset/1080-0.txt': 14, '/content/dataset/pg2600.txt': 169, '/content/dataset/70815.txt.utf-8': 28}, 'like': {'/content/dataset/1998.txt.utf-8': 303, '/content/dataset/pg1513.txt': 39, '/content/dataset/pg11.txt': 85, '/content/dataset/2160-0.txt': 162, '/content/dataset/345.txt.utf-8': 292, '/content/dataset/64317.txt.utf-8': 122, '/content/dataset/70817.txt.utf-8': 42, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 751, '/content/dataset/70815.txt.utf-8': 69}, 'right': {'/content/dataset/1998.txt.utf-8': 58, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 36, '/content/dataset/2160-0.txt': 46, '/content/dataset/345.txt.utf-8': 103, '/content/dataset/64317.txt.utf-8': 54, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 454, '/content/dataset/70815.txt.utf-8': 49}, 'whom': {'/content/dataset/1998.txt.utf-8': 100, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 136, '/content/dataset/345.txt.utf-8': 44, '/content/dataset/64317.txt.utf-8': 10, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 407, '/content/dataset/70815.txt.utf-8': 23}, 'measure': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 25, '/content/dataset/345.txt.utf-8': 11, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 9, '/content/dataset/70815.txt.utf-8': 3}, 'practically': {'/content/dataset/1998.txt.utf-8': 7, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 10}, 'least': {'/content/dataset/1998.txt.utf-8': 22, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 66, '/content/dataset/345.txt.utf-8': 39, '/content/dataset/64317.txt.utf-8': 15, '/content/dataset/70817.txt.ut

l-8': 10, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 85, '/content/dataset/70815.txt.utf-8': 21}, 'when': {'/content/dataset/1998.txt.utf-8': 395, '/content/dataset/pg1513.txt': 54, '/content/dataset/pg11.txt': 80, '/content/dataset/2160-0.txt': 341, '/content/dataset/345.txt.utf-8': 777, '/content/dataset/64317.txt.utf-8': 147, '/content/dataset/70817.txt.utf-8': 80, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 1497, '/content/dataset/70815.txt.utf-8': 220}, 'past': {'/content/dataset/1998.txt.utf-8': 32, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 35, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 16, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 175, '/content/dataset/70815.txt.utf-8': 10}, 'earth': {'/content/dataset/1998.txt.utf-8': 124, '/content/dataset/pg1513.txt': 20, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 30, '/content/dataset/345.txt.utf-8': 50, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 59, '/content/dataset/70815.txt.utf-8': 40}, 'set': {'/content/dataset/1998.txt.utf-8': 34, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 22, '/content/dataset/2160-0.txt': 78, '/content/dataset/345.txt.utf-8': 72, '/content/dataset/64317.txt.utf-8': 24, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 166, '/content/dataset/70815.txt.utf-8': 28}, 'up': {'/content/dataset/1998.txt.utf-8': 189, '/content/dataset/pg1513.txt': 60, '/content/dataset/pg11.txt': 103, '/content/dataset/2160-0.txt': 239, '/content/dataset/345.txt.utf-8': 451, '/content/dataset/64317.txt.utf-8': 194, '/content/dataset/70817.txt.utf-8': 79, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 1583, '/content/dataset/70815.txt.utf-8': 99}, 'much': {'/content/dataset/1998.txt.utf-8': 146, '/content/dataset/pg1513.txt': 38, '/content/dataset/pg11.txt': 52, '/content/dataset/2160-0.txt': 236, '/content/dataset/345.txt.utf-8': 221, '/content/dataset/64317.txt.utf-8': 39, '/content/dataset/70817.txt.utf-8': 19, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 395, '/content/dataset/70815.txt.utf-8': 59}, 'come': {'/content/dataset/1998.txt.utf-8': 140, '/content/dataset/pg1513.txt': 96, '/content/dataset/pg11.txt': 47, '/content/dataset/2160-0.txt': 76, '/content/dataset/345.txt.utf-8': 340, '/content/dataset/64317.txt.utf-8': 85, '/content/dataset/70817.txt.utf-8': 12, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 684, '/content/dataset/70815.txt.utf-8': 112}, 'month': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 9, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 34, '/content/dataset/70815.txt.utf-8': 1}, 'written': {'/content/dataset/1998.txt.utf-8': 23, '/content/dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 17, '/content/dataset/345.txt.utf-8': 35, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 66, '/content/dataset/70815.txt.utf-8': 4}, 'paper': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 20, '/content/dataset/345.txt.utf-8': 27, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 59, '/content/dataset/70815.txt.utf-8': 1}, 'lake': {'/content/dataset/1998.txt.utf-8': 7, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 9, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 12}, 'far': {'/content/dataset/1998.txt.utf-8': 57, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 78, '/content/dataset/345.txt.utf-8': 97, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 21, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 174, '/content/dataset/70815.txt.utf-8': 84}, 'then': {'/content/dataset/1998.txt.utf-8': 357, '/content/dataset/pg1513.txt': 84, '/content/dataset/pg11.txt': 93, '/content/dataset/2160-0.txt': 158, '/content/dataset/345.txt.utf-8': 549, '/content/dataset/64317.txt.utf-8': 144, '/content/dataset/70817.txt.utf-8': 55, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 941, '/content/dataset/70815.txt.utf-8': 196}, 'find': {'/content/dataset/1998.txt.utf-8': 74, '/content/dataset/pg1513.txt': 24, '/content/dataset/pg11.txt': 21, '/content/dataset/2160-0.txt': 113, '/content/dataset/345.txt.utf-8': 115, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 162, '/content/dataset/70815.txt.utf-8': 57}, 'previous': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 39, '/content/dataset/70815.txt.utf-8': 10}

15.txt.utf-8': 2}, 'alteration': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'particularly': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 88, '/content/dataset/70815.txt.utf-8': 10}, 'production': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 2}, 'myself': {'/content/dataset/1998.txt.utf-8': 97, '/content/dataset/pg1513.txt': 15, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 91, '/content/dataset/345.txt.utf-8': 134, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 12, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 148, '/content/dataset/70815.txt.utf-8': 4}, 'small': {'/content/dataset/1998.txt.utf-8': 81, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 55, '/content/dataset/345.txt.utf-8': 43, '/content/dataset/64317.txt.utf-8': 36, '/content/dataset/70817.txt.utf-8': 16, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 181, '/content/dataset/70815.txt.utf-8': 34}, 'discovered': {'/content/dataset/1998.txt.utf-8': 12, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 22, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 14, '/content/dataset/70815.txt.utf-8': 31}, 'page': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 14, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 16, '/content/dataset/70815.txt.utf-8': 3}, 'ten': {'/content/dataset/1998.txt.utf-8': 20, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 24, '/content/dataset/345.txt.utf-8': 14, '/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 100, '/content/dataset/70815.txt.utf-8': 15}, 'your': {'/content/dataset/1998.txt.utf-8': 363, '/content/dataset/pg1513.txt': 113, '/content/dataset/pg11.txt': 64, '/content/dataset/2160-0.txt': 415, '/content/dataset/345.txt.utf-8': 312, '/content/dataset/64317.txt.utf-8': 78, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 820, '/content/dataset/70815.txt.utf-8': 75}, 'many': {'/content/dataset/1998.txt.utf-8': 185, '/content/dataset/pg1513.txt': 17, '/content/dataset/pg11.txt': 14, '/content/dataset/2160-0.txt': 99, '/content/dataset/345.txt.utf-8': 106, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 11, '/content/dataset/pg2600.txt': 235, '/content/dataset/70815.txt.utf-8': 66}, 'beginning': {'/content/dataset/1998.txt.utf-8': 15, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 15, '/content/dataset/2160-0.txt': 11, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 89, '/content/dataset/70815.txt.utf-8': 4}, 'something': {'/content/dataset/1998.txt.utf-8': 65, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 17, '/content/dataset/2160-0.txt': 52, '/content/dataset/345.txt.utf-8': 142, '/content/dataset/64317.txt.utf-8': 68, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 601, '/content/dataset/70815.txt.utf-8': 45}, 'able': {'/content/dataset/1998.txt.utf-8': 19, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 31, '/content/dataset/345.txt.utf-8': 47, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 107, '/content/dataset/70815.txt.utf-8': 18}, 'ever': {'/content/dataset/1998.txt.utf-8': 147, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 20, '/content/dataset/2160-0.txt': 135, '/content/dataset/345.txt.utf-8': 136, '/content/dataset/64317.txt.utf-8': 28, '/content/dataset/70817.txt.utf-8': 14, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 146, '/content/dataset/70815.txt.utf-8': 54}, 'follow': {'/content/dataset/1998.txt.utf-8': 14, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 15, '/content/dataset/345.txt.utf-8': 39, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 40, '/content/dataset/70815.txt.utf-8': 30}, 'created': {'/content/dataset/1998.txt.utf-8': 28, '/content/dataset/pg1513.txt':

2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 7, '/content/dataset/70815.txt.utf-8': 5}, 'bear': {'/content/dataset/1998.txt.utf-8': 23, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 4, 2, '/content/dataset/345.txt.utf-8': 24, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 55, '/content/dataset/70815.txt.utf-8': 25}, 'about': {'/content/dataset/1998.txt.utf-8': 133, '/content/dataset/pg1513.txt': 20, '/content/dataset/pg11.txt': 102, '/content/dataset/2160-0.txt': 152, '/content/dataset/345.txt.utf-8': 240, '/content/dataset/64317.txt.utf-8': 159, '/content/dataset/70817.txt.utf-8': 58, '/content/dataset/1080-0.txt': 13, '/content/dataset/pg2600.txt': 1019, '/content/dataset/70815.txt.utf-8': 133}, 'little': {'/content/dataset/1998.txt.utf-8': 111, '/content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 127, '/content/dataset/2160-0.txt': 148, '/content/dataset/345.txt.utf-8': 168, '/content/dataset/64317.txt.utf-8': 103, '/content/dataset/70817.txt.utf-8': 26, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 557, '/content/dataset/70815.txt.utf-8': 65}, 'cold': {'/content/dataset/1998.txt.utf-8': 34, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 33, '/content/dataset/345.txt.utf-8': 45, '/content/dataset/64317.txt.utf-8': 14, '/content/dataset/70817.txt.utf-8': 22, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 138, '/content/dataset/70815.txt.utf-8': 13}, 'these': {'/content/dataset/1998.txt.utf-8': 135, '/content/dataset/pg1513.txt': 45, '/content/dataset/pg11.txt': 16, '/content/dataset/2160-0.txt': 207, '/content/dataset/345.txt.utf-8': 119, '/content/dataset/64317.txt.utf-8': 28, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 593, '/content/dataset/70815.txt.utf-8': 102}, 'used': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 16, '/content/dataset/2160-0.txt': 38, '/content/dataset/345.txt.utf-8': 46, '/content/dataset/64317.txt.utf-8': 20, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 135, '/content/dataset/70815.txt.utf-8': 13}, 'permitted': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 5, '/content/dataset/70815.txt.utf-8': 8}, 'rather': {'/content/dataset/1998.txt.utf-8': 42, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 25, '/content/dataset/2160-0.txt': 43, '/content/dataset/345.txt.utf-8': 25, '/content/dataset/64317.txt.utf-8': 25, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 88, '/content/dataset/70815.txt.utf-8': 15}, 'walks': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 3, '/content/dataset/70815.txt.utf-8': 1}, 'middle': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 84, '/content/dataset/70815.txt.utf-8': 4}, 'mean': {'/content/dataset/1998.txt.utf-8': 12, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 48, '/content/dataset/345.txt.utf-8': 44, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 66, '/content/dataset/70815.txt.utf-8': 8}, 'matter': {'/content/dataset/1998.txt.utf-8': 40, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 52, '/content/dataset/345.txt.utf-8': 55, '/content/dataset/64317.txt.utf-8': 22, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 173, '/content/dataset/70815.txt.utf-8': 17}, 'met': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 40, '/content/dataset/345.txt.utf-8': 30, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 158, '/content/dataset/70815.txt.utf-8': 10}, 'copies': {'/content/dataset/1998.txt.utf-8': 10, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 7, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 9, '/content/dataset/70815.txt.utf-8': 8}, 'given': {'/content/dataset/1998.txt.utf-8': 37, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 55, '/content/dataset/345.txt.utf-8': 43, '/co

dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 180, '/content/dataset/70815.txt.utf-8': 15}, 'while': {'/content/dataset/1998.txt.utf-8': 41, '/content/dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 26, '/content/dataset/2160-0.txt': 101, '/content/dataset/345.txt.utf-8': 50, '/content/dataset/64317.txt.utf-8': 50, '/content/dataset/70817.txt.utf-8': 15, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 400, '/content/dataset/70815.txt.utf-8': 127}, 'easy': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 19, '/content/dataset/345.txt.utf-8': 16, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 49, '/content/dataset/70815.txt.utf-8': 7}, 'city': {'/content/dataset/1998.txt.utf-8': 19, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 28, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 20, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 67, '/content/dataset/70815.txt.utf-8': 14}, 'leave': {'/content/dataset/1998.txt.utf-8': 18, '/content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 60, '/content/dataset/345.txt.utf-8': 56, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 210, '/content/dataset/70815.txt.utf-8': 10}, 'person': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 116, '/content/dataset/345.txt.utf-8': 15, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 77, '/content/dataset/70815.txt.utf-8': 7}, 'provide': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 7, '/content/dataset/2160-0.txt': 20, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 7}, 'hear': {'/content/dataset/1998.txt.utf-8': 59, '/content/dataset/pg1513.txt': 26, '/content/dataset/pg11.txt': 15, '/content/dataset/2160-0.txt': 39, '/content/dataset/345.txt.utf-8': 102, '/content/dataset/64317.txt.utf-8': 24, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 141, '/content/dataset/70815.txt.utf-8': 8}, 'below': {'/content/dataset/1998.txt.utf-8': 12, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 19, '/content/dataset/345.txt.utf-8': 25, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 46, '/content/dataset/70815.txt.utf-8': 62}, 'increasing': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 20, '/content/dataset/70815.txt.utf-8': 5}, 'describe': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 31, '/content/dataset/70815.txt.utf-8': 2}, 'trouble': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 27, '/content/dataset/345.txt.utf-8': 53, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 33, '/content/dataset/70815.txt.utf-8': 1}, 'allow': {'/content/dataset/1998.txt.utf-8': 7, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 11, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 57, '/content/dataset/70815.txt.utf-8': 1}, 'think': {'/content/dataset/1998.txt.utf-8': 50, '/content/dataset/pg1513.txt': 21, '/content/dataset/pg11.txt': 49, '/content/dataset/2160-0.txt': 128, '/content/dataset/345.txt.utf-8': 218, '/content/dataset/64317.txt.utf-8': 65, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 358, '/content/dataset/70815.txt.utf-8': 32}, 'rest': {'/content/dataset/1998.txt.utf-8': 17, '/content/dataset/pg1513.txt': 16, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 30, '/content/dataset/345.txt.utf-8': 97, '/content/dataset/64317.txt.utf-8': 11, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 108, '/content/dataset/70815.txt.utf-8': 18}, 'neither': {'/content/dataset/1998.txt.utf-8': 27, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 42, '/content/dataset/345.txt.utf-8': 15, '/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 5, '/content/d

dataset/pg2600.txt': 97, '/content/dataset/70815.txt.utf-8': 16}, 'required': {'content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.tx
t': 1, '/content/dataset/2160-0.txt': 10, '/content/dataset/345.txt.utf-8': 10, '/conten
t/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/
1080-0.txt': 1, '/content/dataset/pg2600.txt': 17, '/content/dataset/70815.txt.utf-8':
3}, 'transcribe': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt':
1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/
345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.ut
f-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/d
ataset/70815.txt.utf-8': 1}, 'till': {'/content/dataset/1998.txt.utf-8': 4, '/content/da
taset/pg1513.txt': 22, '/content/dataset/pg11.txt': 21, '/content/dataset/2160-0.txt': 8
5, '/content/dataset/345.txt.utf-8': 135, '/content/dataset/64317.txt.utf-8': 7, '/conte
nt/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg26
00.txt': 174, '/content/dataset/70815.txt.utf-8': 1}, 'age': {'/content/dataset/1998.tx
t.utf-8': 10, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 4, '/conten
t/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 13, '/content/dataset/6431
7.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt':
4, '/content/dataset/pg2600.txt': 44, '/content/dataset/70815.txt.utf-8': 18}, 'mediu
m': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 5, '/content/d
ataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8':
9, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 5, '/cont
ent/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.t
xt.utf-8': 5}, 'does': {'/content/dataset/1998.txt.utf-8': 19, '/content/dataset/pg1513.
txt': 2, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 33, '/content/d
ataset/345.txt.utf-8': 74, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/708
17.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 208,
'/content/dataset/70815.txt.utf-8': 17}, 'takes': {'/content/dataset/1998.txt.utf-8':
5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 3, '/content/dataset/
2160-0.txt': 11, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-
8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/conten
t/dataset/pg2600.txt': 26, '/content/dataset/70815.txt.utf-8': 2}, 'ask': {'/content/dat
aset/1998.txt.utf-8': 26, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt':
11, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 70, '/content/d
ataset/64317.txt.utf-8': 15, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/10
80-0.txt': 1, '/content/dataset/pg2600.txt': 200, '/content/dataset/70815.txt.utf-8':
8}, 'either': {'/content/dataset/1998.txt.utf-8': 17, '/content/dataset/pg1513.txt': 1
0, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 53, '/content/dase
t/345.txt.utf-8': 39, '/content/dataset/64317.txt.utf-8': 14, '/content/dataset/70817.tx
t.utf-8': 4, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 128, '/con
tent/dataset/70815.txt.utf-8': 41}, 'fine': {'/content/dataset/1998.txt.utf-8': 18, '/co
nrent/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.t
xt': 49, '/content/dataset/345.txt.utf-8': 21, '/content/dataset/64317.txt.utf-8': 11,
'/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/data
set/pg2600.txt': 122, '/content/dataset/70815.txt.utf-8': 1}, 'need': {'/content/dase
t/1998.txt.utf-8': 22, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt':
1, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 69, '/content/d
ataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/108
0-0.txt': 1, '/content/dataset/pg2600.txt': 108, '/content/dataset/70815.txt.utf-8': 1
4}, 'sort': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/c
ontent/dataset/pg11.txt': 20, '/content/dataset/2160-0.txt': 28, '/content/dataset/345.t
xt.utf-8': 81, '/content/dataset/64317.txt.utf-8': 23, '/content/dataset/70817.txt.utf-
8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 58, '/content/da
taset/70815.txt.utf-8': 7}, 'own': {'/content/dataset/1998.txt.utf-8': 128, '/content/da
taset/pg1513.txt': 17, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 3
28, '/content/dataset/345.txt.utf-8': 187, '/content/dataset/64317.txt.utf-8': 31, '/con
tent/dataset/70817.txt.utf-8': 26, '/content/dataset/1080-0.txt': 10, '/content/dataset/
pg2600.txt': 502, '/content/dataset/70815.txt.utf-8': 64}, 'want': {'/content/dataset/19
98.txt.utf-8': 89, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 9, '/c
ontent/dataset/2160-0.txt': 51, '/content/dataset/345.txt.utf-8': 116, '/content/dase
t/64317.txt.utf-8': 59, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.
txt': 5, '/content/dataset/pg2600.txt': 251, '/content/dataset/70815.txt.utf-8': 17}, 'u
pon': {'/content/dataset/1998.txt.utf-8': 144, '/content/dataset/pg1513.txt': 44, '/cont
ent/dataset/pg11.txt': 28, '/content/dataset/2160-0.txt': 466, '/content/dataset/345.tx
t.utf-8': 90, '/content/dataset/64317.txt.utf-8': 35, '/content/dataset/70817.txt.utf-
8': 11, '/content/dataset/1080-0.txt': 15, '/content/dataset/pg2600.txt': 97, '/content/
dataset/70815.txt.utf-8': 483}, 'learn': {'/content/dataset/1998.txt.utf-8': 83, '/conte

nt/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 36, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 18, '/content/dataset/70815.txt.utf-8': 6}, 'doubt': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 56, '/content/dataset/345.txt.utf-8': 41, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 60, '/content/dataset/70815.txt.utf-8': 7}, 'few': {'/content/dataset/1998.txt.utf-8': 12, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 69, '/content/dataset/345.txt.utf-8': 86, '/content/dataset/64317.txt.utf-8': 26, '/content/dataset/70817.txt.utf-8': 25, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 171, '/content/dataset/70815.txt.utf-8': 44}, 'looked': {'/content/dataset/1998.txt.utf-8': 37, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 45, '/content/dataset/2160-0.txt': 38, '/content/dataset/345.txt.utf-8': 186, '/content/dataset/64317.txt.utf-8': 85, '/content/dataset/70817.txt.utf-8': 33, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 651, '/content/dataset/70815.txt.utf-8': 58}, 'let': {'/content/dataset/1998.txt.utf-8': 108, '/content/dataset/pg1513.txt': 50, '/content/dataset/pg11.txt': 22, '/content/dataset/2160-0.txt': 55, '/content/dataset/345.txt.utf-8': 148, '/content/dataset/64317.txt.utf-8': 38, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 405, '/content/dataset/70815.txt.utf-8': 41}, 'without': {'/content/dataset/1998.txt.utf-8': 73, '/content/dataset/pg1513.txt': 22, '/content/dataset/pg11.txt': 34, '/content/dataset/2160-0.txt': 307, '/content/dataset/345.txt.utf-8': 143, '/content/dataset/64317.txt.utf-8': 35, '/content/dataset/70817.txt.utf-8': 31, '/content/dataset/1080-0.txt': 12, '/content/dataset/pg2600.txt': 675, '/content/dataset/70815.txt.utf-8': 57}, 'walk': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 17, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 47, '/content/dataset/70815.txt.utf-8': 6}, 'less': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 46, '/content/dataset/345.txt.utf-8': 31, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 143, '/content/dataset/70815.txt.utf-8': 20}, 'alone': {'/content/dataset/1998.txt.utf-8': 62, '/content/dataset/pg1513.txt': 16, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 10, '/content/dataset/345.txt.utf-8': 81, '/content/dataset/64317.txt.utf-8': 34, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 243, '/content/dataset/70815.txt.utf-8': 37}, 'printed': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 9, '/content/dataset/70815.txt.utf-8': 3}, 'original': {'/content/dataset/1998.txt.utf-8': 10, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 16, '/content/dataset/70815.txt.utf-8': 2}, 'public': {'/content/dataset/1998.txt.utf-8': 11, '/content/dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 63, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 28, '/content/dataset/70815.txt.utf-8': 4}, 'making': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 42, '/content/dataset/345.txt.utf-8': 31, '/content/dataset/64317.txt.utf-8': 11, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 109, '/content/dataset/70815.txt.utf-8': 9}, 'distribute': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 7, '/content/dataset/70815.txt.utf-8': 8}, 'forty': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 18, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 46, '/content/dataset/70815.txt.utf-8': 2}, 'others': {'/content/dataset/1998.txt.utf-8': 35, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 21, '/content/dataset/345.txt.utf-8': 82, '/co

```
'content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 303, '/content/dataset/70815.txt.utf-8': 49}, 'name': {'/content/dataset/1998.txt.utf-8': 26, '/content/dataset/pg1513.txt': 27, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 88, '/content/dataset/345.txt.utf-8': 21, '/content/dataset/64317.txt.utf-8': 43, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 120, '/content/dataset/70815.txt.utf-8': 13}, 'see': {'/content/dataset/1998.txt.utf-8': 154, '/content/dataset/pg1513.txt': 40, '/content/dataset/pg11.txt': 70, '/content/dataset/2160-0.txt': 117, '/content/dataset/345.txt.utf-8': 398, '/content/dataset/64317.txt.utf-8': 84, '/content/dataset/70817.txt.utf-8': 31, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 734, '/content/dataset/70815.txt.utf-8': 69}, 'cause': {'/content/dataset/1998.txt.utf-8': 21, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 31, '/content/dataset/345.txt.utf-8': 51, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 148, '/content/dataset/70815.txt.utf-8': 12}, 'answer': {'/content/dataset/1998.txt.utf-8': 14, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 21, '/content/dataset/345.txt.utf-8': 44, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 143, '/content/dataset/70815.txt.utf-8': 8}, 'heart': {'/content/dataset/1998.txt.utf-8': 185, '/content/dataset/pg1513.txt': 33, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 98, '/content/dataset/345.txt.utf-8': 141, '/content/dataset/64317.txt.utf-8': 19, '/content/dataset/70817.txt.utf-8': 15, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 179, '/content/dataset/70815.txt.utf-8': 20}, 'poor': {'/content/dataset/1998.txt.utf-8': 29, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 27, '/content/dataset/2160-0.txt': 168, '/content/dataset/345.txt.utf-8': 193, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 7, '/content/dataset/pg2600.txt': 77, '/content/dataset/70815.txt.utf-8': 1}, 'carry': {'/content/dataset/1998.txt.utf-8': 21, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 17, '/content/dataset/345.txt.utf-8': 16, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 53, '/content/dataset/70815.txt.utf-8': 16}, 'ago': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 47, '/content/dataset/345.txt.utf-8': 26, '/content/dataset/64317.txt.utf-8': 19, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 74, '/content/dataset/70815.txt.utf-8': 6}, 'said': {'/content/dataset/1998.txt.utf-8': 199, '/content/dataset/pg1513.txt': 14, '/content/dataset/pg11.txt': 460, '/content/dataset/2160-0.txt': 280, '/content/dataset/345.txt.utf-8': 570, '/content/dataset/64317.txt.utf-8': 235, '/content/dataset/70817.txt.utf-8': 37, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 2842, '/content/dataset/70815.txt.utf-8': 386}, 'accept': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 24, '/content/dataset/70815.txt.utf-8': 6}, 'believe': {'/content/dataset/1998.txt.utf-8': 52, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 74, '/content/dataset/345.txt.utf-8': 75, '/content/dataset/64317.txt.utf-8': 16, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 124, '/content/dataset/70815.txt.utf-8': 27}, 'stay': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 30, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 26, '/content/dataset/345.txt.utf-8': 43, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 57, '/content/dataset/70815.txt.utf-8': 7}, 'make': {'/content/dataset/1998.txt.utf-8': 77, '/content/dataset/pg1513.txt': 43, '/content/dataset/pg11.txt': 30, '/content/dataset/2160-0.txt': 192, '/content/dataset/345.txt.utf-8': 165, '/content/dataset/64317.txt.utf-8': 30, '/content/dataset/70817.txt.utf-8': 14, '/content/dataset/1080-0.txt': 11, '/content/dataset/pg2600.txt': 278, '/content/dataset/70815.txt.utf-8': 43}, 'bring': {'/content/dataset/1998.txt.utf-8': 13, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 26, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 83, '/content/dataset/70815.txt.utf-8': 3}, 'same': {'/content/dataset/1998.txt.utf-8': 40, '/content/dataset/pg1513.txt': 16, '/content/dataset/pg11.txt': 25, '/content/dataset/2160-0.txt': 143, '/content/dataset/345.txt.utf-8': 111, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.tx
```


't': 2, '/content/dataset/pg2600.txt': 665, '/content/dataset/70815.txt.utf-8': 54, 'remain': {'/content/dataset/1998.txt.utf-8': 14, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 23, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 77, '/content/dataset/70815.txt.utf-8': 23}, 'true': {'/content/dataset/1998.txt.utf-8': 32, '/content/dataset/pg1513.txt': 28, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 24, '/content/dataset/345.txt.utf-8': 58, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 120, '/content/dataset/70815.txt.utf-8': 14}, 'against': {'/content/dataset/1998.txt.utf-8': 73, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 60, '/content/dataset/345.txt.utf-8': 68, '/content/dataset/64317.txt.utf-8': 33, '/content/dataset/70817.txt.utf-8': 20, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 224, '/content/dataset/70815.txt.utf-8': 35}, 'receive': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 24, '/content/dataset/345.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 43, '/content/dataset/70815.txt.utf-8': 5}, 'reason': {'/content/dataset/1998.txt.utf-8': 26, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 44, '/content/dataset/345.txt.utf-8': 28, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 126, '/content/dataset/70815.txt.utf-8': 20}, 'except': {'/content/dataset/1998.txt.utf-8': 23, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 31, '/content/dataset/345.txt.utf-8': 41, '/content/dataset/64317.txt.utf-8': 21, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 73, '/content/dataset/70815.txt.utf-8': 21}, 'house': {'/content/dataset/1998.txt.utf-8': 17, '/content/dataset/pg1513.txt': 33, '/content/dataset/pg11.txt': 18, '/content/dataset/2160-0.txt': 221, '/content/dataset/345.txt.utf-8': 135, '/content/dataset/64317.txt.utf-8': 95, '/content/dataset/70817.txt.utf-8': 61, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 397, '/content/dataset/70815.txt.utf-8': 1}, 'prepare': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 11, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 29, '/content/dataset/70815.txt.utf-8': 4}, 'share': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 17, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 29, '/content/dataset/70815.txt.utf-8': 4}, 'keep': {'/content/dataset/1998.txt.utf-8': 30, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 59, '/content/dataset/345.txt.utf-8': 98, '/content/dataset/64317.txt.utf-8': 13, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 84, '/content/dataset/70815.txt.utf-8': 12}, 'free': {'/content/dataset/1998.txt.utf-8': 43, '/content/dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 42, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 156, '/content/dataset/70815.txt.utf-8': 23}, 'head': {'/content/dataset/1998.txt.utf-8': 54, '/content/dataset/pg1513.txt': 16, '/content/dataset/pg11.txt': 50, '/content/dataset/2160-0.txt': 92, '/content/dataset/345.txt.utf-8': 106, '/content/dataset/64317.txt.utf-8': 40, '/content/dataset/70817.txt.utf-8': 26, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 568, '/content/dataset/70815.txt.utf-8': 80}, 'stand': {'/content/dataset/1998.txt.utf-8': 57, '/content/dataset/pg1513.txt': 26, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 15, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 52, '/content/dataset/70815.txt.utf-8': 9}, 'fix': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 1}, 'grow': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 14, '/content/dataset/345.txt.utf-8': 21, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 28, '/content/dataset/70815.txt.utf-8': 2}, 'soon': {'/content/dataset/1998.txt.utf-8': 20, '/co

content/dataset/pg1513.txt': 14, '/content/dataset/pg11.txt': 25, '/content/dataset/2160-0.txt': 40, '/content/dataset/345.txt.utf-8': 68, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 252, '/content/dataset/70815.txt.utf-8': 27}, 'course': {'/content/dataset/1998.txt.utf-8': 23, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 27, '/content/dataset/2160-0.txt': 45, '/content/dataset/345.txt.utf-8': 42, '/content/dataset/64317.txt.utf-8': 23, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 124, '/content/dataset/70815.txt.utf-8': 35}, 'better': {'/content/dataset/1998.txt.utf-8': 61, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 14, '/content/dataset/2160-0.txt': 47, '/content/dataset/345.txt.utf-8': 77, '/content/dataset/64317.txt.utf-8': 18, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 163, '/content/dataset/70815.txt.utf-8': 29}, 'arms': {'/content/dataset/1998.txt.utf-8': 7, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 27, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 13, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 176, '/content/dataset/70815.txt.utf-8': 11}, 'especially': {'/content/dataset/1998.txt.utf-8': 32, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 17, '/content/dataset/345.txt.utf-8': 10, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 140, '/content/dataset/70815.txt.utf-8': 7}, 'fell': {'/content/dataset/1998.txt.utf-8': 12, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 59, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 125, '/content/dataset/70815.txt.utf-8': 25}, 'knew': {'/content/dataset/1998.txt.utf-8': 16, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 15, '/content/dataset/2160-0.txt': 37, '/content/dataset/345.txt.utf-8': 128, '/content/dataset/64317.txt.utf-8': 54, '/content/dataset/70817.txt.utf-8': 21, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 413, '/content/dataset/70815.txt.utf-8': 89}, 'prevent': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 11, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 26, '/content/dataset/70815.txt.utf-8': 3}, 'further': {'/content/dataset/1998.txt.utf-8': 22, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 41, '/content/dataset/345.txt.utf-8': 34, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 51, '/content/dataset/70815.txt.utf-8': 21}, 'ways': {'/content/dataset/1998.txt.utf-8': 14, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 9, '/content/dataset/345.txt.utf-8': 21, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 16, '/content/dataset/70815.txt.utf-8': 5}, 'gone': {'/content/dataset/1998.txt.utf-8': 33, '/content/dataset/pg1513.txt': 33, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 31, '/content/dataset/345.txt.utf-8': 73, '/content/dataset/64317.txt.utf-8': 23, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 162, '/content/dataset/70815.txt.utf-8': 36}, 'hundred': {'/content/dataset/1998.txt.utf-8': 15, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 58, '/content/dataset/345.txt.utf-8': 12, '/content/dataset/64317.txt.utf-8': 14, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 122, '/content/dataset/70815.txt.utf-8': 33}, 'late': {'/content/dataset/1998.txt.utf-8': 13, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 13, '/content/dataset/345.txt.utf-8': 60, '/content/dataset/64317.txt.utf-8': 13, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 106, '/content/dataset/70815.txt.utf-8': 2}, 'immediately': {'/content/dataset/1998.txt.utf-8': 15, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 56, '/content/dataset/345.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 16, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 130, '/content/dataset/70815.txt.utf-8': 17}, 'offer': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 36, '/content/dataset/70815.txt.utf-8': 2}, 'protect': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 7, '/content

'dataset/345.txt.utf-8': 13, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 9}, 'law': {'/content/dataset/1998.txt.utf-8': 33, '/content/dataset/pg1513.txt': 19, '/content/dataset/pg11.txt': 9, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 21, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt': 135, '/content/dataset/70815.txt.utf-8': 9}, 'disclaim': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'hot': {'/content/dataset/1998.txt.utf-8': 15, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 7, '/content/dataset/2160-0.txt': 37, '/content/dataset/345.txt.utf-8': 11, '/content/dataset/64317.txt.utf-8': 25, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 49, '/content/dataset/70815.txt.utf-8': 2}, 'cannot': {'/content/dataset/1998.txt.utf-8': 35, '/content/dataset/pg1513.txt': 25, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 102, '/content/dataset/345.txt.utf-8': 90, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 10, '/content/dataset/pg2600.txt': 168, '/content/dataset/70815.txt.utf-8': 36}, 'times': {'/content/dataset/1998.txt.utf-8': 25, '/content/dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 45, '/content/dataset/345.txt.utf-8': 56, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 154, '/content/dataset/70815.txt.utf-8': 22}, 'send': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 14, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 70, '/content/dataset/70815.txt.utf-8': 6}, 'best': {'/content/dataset/1998.txt.utf-8': 101, '/content/dataset/pg1513.txt': 12, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 60, '/content/dataset/345.txt.utf-8': 51, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 128, '/content/dataset/70815.txt.utf-8': 17}, 'please': {'/content/dataset/1998.txt.utf-8': 11, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 21, '/content/dataset/2160-0.txt': 23, '/content/dataset/345.txt.utf-8': 24, '/content/dataset/64317.txt.utf-8': 13, '/content/dataset/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 138, '/content/dataset/70815.txt.utf-8': 5}, 'choose': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 9, '/content/dataset/345.txt.utf-8': 8, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 26, '/content/dataset/70815.txt.utf-8': 4}, 'sure': {'/content/dataset/1998.txt.utf-8': 25, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg11.txt': 24, '/content/dataset/2160-0.txt': 68, '/content/dataset/345.txt.utf-8': 47, '/content/dataset/64317.txt.utf-8': 17, '/content/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 69, '/content/dataset/70815.txt.utf-8': 22}, 'get': {'/content/dataset/1998.txt.utf-8': 19, '/content/dataset/pg1513.txt': 13, '/content/dataset/pg11.txt': 46, '/content/dataset/2160-0.txt': 26, '/content/dataset/345.txt.utf-8': 172, '/content/dataset/64317.txt.utf-8': 75, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 357, '/content/dataset/70815.txt.utf-8': 19}, 'freely': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 17, '/content/dataset/70815.txt.utf-8': 5}, 'indeed': {'/content/dataset/1998.txt.utf-8': 31, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 16, '/content/dataset/2160-0.txt': 95, '/content/dataset/345.txt.utf-8': 52, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 33, '/content/dataset/70815.txt.utf-8': 13}, 'wish': {'/content/dataset/1998.txt.utf-8': 25, '/content/dataset/pg1513.txt': 7, '/content/dataset/pg11.txt': 22, '/content/dataset/2160-0.txt': 29, '/content/dataset/345.txt.utf-8': 51, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 188, '/content/dataset/70815.txt.utf-8': 16}, 'children': {'/content/dataset/1998.txt.utf-8': 60, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 10, '/content/dataset/2160-0.txt': 15, '/content/dataset/345.txt.utf-8': 22, '/content/dataset/64317.txt.utf-8': 7, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1}

117, '/content/dataset/pg2600.txt': 117, '/content/dataset/70815.txt.utf-8': 11, 'die': {'/content/dataset/1998.txt.utf-8': 32, '/content/dataset/pg1513.txt': 23, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 45, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 68, '/content/dataset/70815.txt.utf-8': 16}, 'though': {'/content/dataset/1998.txt.utf-8': 30, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 131, '/content/dataset/345.txt.utf-8': 221, '/content/dataset/64317.txt.utf-8': 35, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 456, '/content/dataset/70815.txt.utf-8': 76}, 'her': {'/content/dataset/1998.txt.utf-8': 37, '/content/dataset/pg1513.txt': 154, '/content/dataset/pg11.txt': 247, '/content/dataset/2160-0.txt': 761, '/content/dataset/345.txt.utf-8': 1059, '/content/dataset/64317.txt.utf-8': 398, '/content/dataset/70817.txt.utf-8': 157, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 4725, '/content/dataset/70815.txt.utf-8': 391}, 'unless': {'/content/dataset/1998.txt.utf-8': 24, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 13, '/content/dataset/345.txt.utf-8': 25, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 15, '/content/dataset/70815.txt.utf-8': 15}, 'lot': {'/content/dataset/1998.txt.utf-8': 8, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 17, '/content/dataset/64317.txt.utf-8': 10, '/content/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 21, '/content/dataset/70815.txt.utf-8': 30}, 'off': {'/content/dataset/1998.txt.utf-8': 17, '/content/dataset/pg1513.txt': 10, '/content/dataset/pg11.txt': 73, '/content/dataset/2160-0.txt': 76, '/content/dataset/345.txt.utf-8': 155, '/content/dataset/64317.txt.utf-8': 45, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 431, '/content/dataset/70815.txt.utf-8': 38}, 'interpreted': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 3, '/content/dataset/70815.txt.utf-8': 3}, 'read': {'/content/dataset/1998.txt.utf-8': 26, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 14, '/content/dataset/2160-0.txt': 26, '/content/dataset/345.txt.utf-8': 84, '/content/dataset/64317.txt.utf-8': 14, '/content/dataset/70817.txt.utf-8': 13, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 133, '/content/dataset/70815.txt.utf-8': 5}, 'she': {'/content/dataset/1998.txt.utf-8': 47, '/content/dataset/pg1513.txt': 125, '/content/dataset/pg11.txt': 549, '/content/dataset/2160-0.txt': 558, '/content/dataset/345.txt.utf-8': 817, '/content/dataset/64317.txt.utf-8': 413, '/content/dataset/70817.txt.utf-8': 290, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 3489, '/content/dataset/70815.txt.utf-8': 334}, 'method': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 12, '/content/dataset/70815.txt.utf-8': 3}, 'since': {'/content/dataset/1998.txt.utf-8': 17, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 51, '/content/dataset/345.txt.utf-8': 86, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 9, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 188, '/content/dataset/70815.txt.utf-8': 63}, 'move': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 9, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 27, '/content/dataset/64317.txt.utf-8': 10, '/content/dataset/70817.txt.utf-8': 10, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 84, '/content/dataset/70815.txt.utf-8': 12}, 'hold': {'/content/dataset/1998.txt.utf-8': 18, '/content/dataset/pg1513.txt': 18, '/content/dataset/pg11.txt': 11, '/content/dataset/2160-0.txt': 34, '/content/dataset/345.txt.utf-8': 32, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 64, '/content/dataset/70815.txt.utf-8': 23}, 'bent': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 26, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 62, '/content/dataset/70815.txt.utf-8': 3}, 'discover': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 9, '/content/dataset/70815.txt.utf-8': 11}, 'change': {'/content/dataset/199

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

asset/64317.txt.utf-8': 4, '/content/dataset/pg2600.txt': 61, '/content/dataset/70815.txt.utf-8': 2},
 'variety': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/c
 ontent/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 23, '/content/dataset/345.tx
 t.utf-8': 1, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8':
 2, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 12, '/content/datas
 et/70815.txt.utf-8': 2}, 'outside': {'/content/dataset/1998.txt.utf-8': 8, '/content/dat
 aset/pg1513.txt': 2, '/content/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 6,
 '/content/dataset/345.txt.utf-8': 37, '/content/dataset/64317.txt.utf-8': 18, '/conten
 t/dataset/70817.txt.utf-8': 18, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg26
 00.txt': 47, '/content/dataset/70815.txt.utf-8': 5}, 'paying': {'/content/dataset/1998.t
 xt.utf-8': 4, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/conten
 t/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 4, '/content/dataset/64317.t
 xt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 4,
 '/content/dataset/pg2600.txt': 17, '/content/dataset/70815.txt.utf-8': 4}, 'domain':
 {'/content/dataset/1998.txt.utf-8': 13, '/content/dataset/pg1513.txt': 2, '/content/dat
 aset/pg11.txt': 2, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8':
 2, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 3, '/cont
 ent/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 7, '/content/dataset/70815.tx
 t.utf-8': 4}, 'nearly': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.
 txt': 2, '/content/dataset/pg11.txt': 13, '/content/dataset/2160-0.txt': 9, '/content/da
 taset/345.txt.utf-8': 22, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/7081
 7.txt.utf-8': 18, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 57,
 '/content/dataset/70815.txt.utf-8': 3}, 'secure': {'/content/dataset/1998.txt.utf-8':
 2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/
 2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8':
 3, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 1, '/content/d
 ataset/pg2600.txt': 16, '/content/dataset/70815.txt.utf-8': 1}, 'imposed': {'/content/da
 taset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt':
 1, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 1, '/content/dat
 aset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-
 0.txt': 1, '/content/dataset/pg2600.txt': 4, '/content/dataset/70815.txt.utf-8': 1}, 'tw
 enty': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 11, '/conte
 nt/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 48, '/content/dataset/345.txt.ut
 f-8': 13, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 4,
 '/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 135, '/content/datase
 t/70815.txt.utf-8': 17}, 'committed': {'/content/dataset/1998.txt.utf-8': 2, '/content/d
 ataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1
 0, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 2, '/conten
 t/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg260
 0.txt': 17, '/content/dataset/70815.txt.utf-8': 1}, 'efforts': {'/content/dataset/1998.t
 xt.utf-8': 4, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/conten
 t/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 12, '/content/dataset/6431
 7.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt':
 3, '/content/dataset/pg2600.txt': 45, '/content/dataset/70815.txt.utf-8': 9}, 'notice':
 {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 4, '/content/data
 set/pg11.txt': 8, '/content/dataset/2160-0.txt': 25, '/content/dataset/345.txt.utf-8': 3
 2, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 3, '/conte
 nt/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 70, '/content/dataset/70815.tx
 t.utf-8': 8}, 'determine': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg15
 13.txt': 3, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/
 dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/7081
 7.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 8, '/c
 ontent/dataset/70815.txt.utf-8': 3}, 'profit': {'/content/dataset/1998.txt.utf-8': 3, '
 /content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-
 0.txt': 6, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 1,
 '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 4, '/content/data
 set/pg2600.txt': 6, '/content/dataset/70815.txt.utf-8': 2}, 'six': {'/content/dataset/19
 98.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 2, '/co
 ntent/dataset/2160-0.txt': 23, '/content/dataset/345.txt.utf-8': 12, '/content/dataset/6
 4317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.tx
 t': 3, '/content/dataset/pg2600.txt': 61, '/content/dataset/70815.txt.utf-8': 17}, 'proh
 ibition': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/con
 tent/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.u
 tf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1,
 '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/7

[illegible]

```
t.txt.utf-8': 11, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.tx
t.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 14, '/cont
ent/dataset/70815.txt.utf-8': 7}, 'assistance': {'/content/dataset/1998.txt.utf-8': 2,
'/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/216
0-0.txt': 26, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8':
1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/d
ataset/pg2600.txt': 8, '/content/dataset/70815.txt.utf-8': 7}, 'c': {'/content/dataset/1
998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 6, '/content/dataset/pg11.txt': 5, '/c
ontent/dataset/2160-0.txt': 34, '/content/dataset/345.txt.utf-8': 7, '/content/dataset/6
4317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.tx
t': 4, '/content/dataset/pg2600.txt': 18, '/content/dataset/70815.txt.utf-8': 4}, 'accep
ted': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 2, '/conten
t/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-
8': 4, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset/70817.txt.utf-8': 4, '/c
ontent/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 36, '/content/dataset/7081
5.txt.utf-8': 11}, 'explanation': {'/content/dataset/1998.txt.utf-8': 4, '/content/datas
et/pg1513.txt': 2, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 11, '/
content/dataset/345.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 2, '/content/data
set/70817.txt.utf-8': 6, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.tx
t': 40, '/content/dataset/70815.txt.utf-8': 6}, 'd': {'/content/dataset/1998.txt.utf-8':
2, '/content/dataset/pg1513.txt': 242, '/content/dataset/pg11.txt': 29, '/content/datas
et/2160-0.txt': 31, '/content/dataset/345.txt.utf-8': 27, '/content/dataset/64317.txt.ut
f-8': 103, '/content/dataset/70817.txt.utf-8': 17, '/content/dataset/1080-0.txt': 2, '/c
ontent/dataset/pg2600.txt': 111, '/content/dataset/70815.txt.utf-8': 3}, 'including':
{'/content/dataset/1998.txt.utf-8': 10, '/content/dataset/pg1513.txt': 9, '/content/dat
aset/pg11.txt': 9, '/content/dataset/2160-0.txt': 17, '/content/dataset/345.txt.utf-8':
10, '/content/dataset/64317.txt.utf-8': 12, '/content/dataset/70817.txt.utf-8': 10, '/c
ontent/dataset/1080-0.txt': 9, '/content/dataset/pg2600.txt': 16, '/content/dataset/7081
5.txt.utf-8': 11}, 'effort': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg
1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 13, '/conte
nt/dataset/345.txt.utf-8': 38, '/content/dataset/64317.txt.utf-8': 15, '/content/dataset
/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt':
78, '/content/dataset/70815.txt.utf-8': 32}, 'paragraph': {'/content/dataset/1998.txt.u
tf-8': 23, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt': 11, '/conten
t/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 12, '/content/dataset/6431
7.txt.utf-8': 11, '/content/dataset/70817.txt.utf-8': 12, '/content/dataset/1080-0.txt':
11, '/content/dataset/pg2600.txt': 13, '/content/dataset/70815.txt.utf-8': 11}, 'make
s': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 13, '/content/
dataset/pg11.txt': 12, '/content/dataset/2160-0.txt': 19, '/content/dataset/345.txt.utf-
8': 20, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/
content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 27, '/content/dataset/708
15.txt.utf-8': 4}, 'despite': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/p
g1513.txt': 2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/conte
nt/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/7
0817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 84,
'/content/dataset/70815.txt.utf-8': 2}, 'reported': {'/content/dataset/1998.txt.utf-8':
2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/
2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 11, '/content/dataset/64317.txt.utf-
8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/conten
t/dataset/pg2600.txt': 36, '/content/dataset/70815.txt.utf-8': 4}, 'f': {'/content/datas
et/1998.txt.utf-8': 12, '/content/dataset/pg1513.txt': 11, '/content/dataset/pg11.txt':
11, '/content/dataset/2160-0.txt': 12, '/content/dataset/345.txt.utf-8': 16, '/content/
dataset/64317.txt.utf-8': 15, '/content/dataset/70817.txt.utf-8': 11, '/content/dataset/
1080-0.txt': 11, '/content/dataset/pg2600.txt': 15, '/content/dataset/70815.txt.utf-8':
11}, 'perform': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt':
2, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 8, '/content/dataset/
345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.ut
f-8': 1, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 16, '/content/
dataset/70815.txt.utf-8': 1}, 'offers': {'/content/dataset/1998.txt.utf-8': 2, '/conten
t/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt':
2, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/conten
t/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg260
0.txt': 7, '/content/dataset/70815.txt.utf-8': 1}, 'appears': {'/content/dataset/1998.tx
t.utf-8': 4, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 1, '/conten
t/dataset/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.
txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1,
```

```

/content/dataset/pg2600.txt': 24, '/content/dataset/pg2600.txt.utf-8': 1}, 'breach':
{'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 2, '/content/data
set/pg11.txt': 2, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 2,
'/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/conten
t/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 3, '/content/dataset/70815.txt.
utf-8': 2}, 'pages': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg1513.tx
t': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/datas
et/345.txt.utf-8': 4, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70817.tx
t.utf-8': 4, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 9, '/conte
nt/dataset/70815.txt.utf-8': 1}, 'addresses': {'/content/dataset/1998.txt.utf-8': 2, '/c
ontent/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.
txt': 10, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 2, '/
content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset
/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'display': {'/content/dataset/
1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/
content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/6
4317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.tx
t': 1, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 2}, 'conce
pt': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 2, '/content/
dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-
8': 2, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/c
ontent/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 3, '/content/dataset/7081
5.txt.utf-8': 2}, 'associated': {'/content/dataset/1998.txt.utf-8': 9, '/content/dataset
/pg1513.txt': 7, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 8, '/co
ntent/dataset/345.txt.utf-8': 7, '/content/dataset/64317.txt.utf-8': 9, '/content/dataset
/70817.txt.utf-8': 7, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg2600.txt':
14, '/content/dataset/70815.txt.utf-8': 8}, 'support': {'/content/dataset/1998.txt.utf-
8': 6, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/datas
et/2160-0.txt': 16, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf
-8': 6, '/content/dataset/70817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 5, '/conte
nt/dataset/pg2600.txt': 20, '/content/dataset/70815.txt.utf-8': 10}, 'direct': {'/conten
t/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.t
xt': 1, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.txt.utf-8': 5, '/conten
t/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/
1080-0.txt': 1, '/content/dataset/pg2600.txt': 49, '/content/dataset/70815.txt.utf-8':
4}, 'generations': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.tx
t': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/datas
et/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.tx
t.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 4, '/conte
nt/dataset/70815.txt.utf-8': 3}, 'extent': {'/content/dataset/1998.txt.utf-8': 4, '/cont
ent/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.tx
t': 8, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 1, '/con
tent/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg
2600.txt': 13, '/content/dataset/70815.txt.utf-8': 7}, 'returns': {'/content/dataset/199
8.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/con
tent/dataset/2160-0.txt': 7, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/6431
7.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt':
1, '/content/dataset/pg2600.txt': 8, '/content/dataset/70815.txt.utf-8': 7}, 'limited':
{'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 5, '/content/data
set/pg11.txt': 5, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 8,
'/content/dataset/64317.txt.utf-8': 8, '/content/dataset/70817.txt.utf-8': 6, '/conten
t/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 18, '/content/dataset/70815.tx
t.utf-8': 6}, 'strict': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.
txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/content/dat
aset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.t
xt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 8, '/cont
ent/dataset/70815.txt.utf-8': 2}, 'indirectly': {'/content/dataset/1998.txt.utf-8': 2,
'/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/216
0-0.txt': 1, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 2,
'/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/data
set/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'edition': {'/content/dataset
/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 2,
'/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset
/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.t
xt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 3}, 'reach
ing': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 1, '/conten

```

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```
content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt'.
utf-8': 3, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 2,
'/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 6, '/content/dataset/7
0815.txt.utf-8': 1}, 'applicable': {'/content/dataset/1998.txt.utf-8': 3, '/content/data
set/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 4, '/
content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 3, '/content/data
set/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.tx
t': 5, '/content/dataset/70815.txt.utf-8': 3}, 'taxes': {'/content/dataset/1998.txt.utf-
8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/datas
et/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-
8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/conten
t/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'owner': {'/content/da
taset/1998.txt.utf-8': 4, '/content/dataset/pg1513.txt': 4, '/content/dataset/pg11.txt':
4, '/content/dataset/2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 10, '/content/da
taset/64317.txt.utf-8': 4, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080
-0.txt': 4, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 4},
'donate': {'/content/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 5, '/co
ntent/dataset/pg11.txt': 5, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.
utf-8': 5, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 5,
'/content/dataset/1080-0.txt': 5, '/content/dataset/pg2600.txt': 5, '/content/dataset/7
0815.txt.utf-8': 5}, 'payments': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset
/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 4, '/co
ntent/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset
/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt':
6, '/content/dataset/70815.txt.utf-8': 3}, 'legally': {'/content/dataset/1998.txt.utf-
8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/datas
et/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-
8': 2, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/conten
t/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'periodic': {'/conten
t/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.t
xt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/conten
t/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/
1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.utf-8':
1}, 'tax': {'/content/dataset/1998.txt.utf-8': 6, '/content/dataset/pg1513.txt': 6, '/c
ontent/dataset/pg11.txt': 6, '/content/dataset/2160-0.txt': 8, '/content/dataset/345.tx
t.utf-8': 9, '/content/dataset/64317.txt.utf-8': 6, '/content/dataset/70817.txt.utf-8':
6, '/content/dataset/1080-0.txt': 6, '/content/dataset/pg2600.txt': 6, '/content/dataset
/70815.txt.utf-8': 6}, 'information': {'/content/dataset/1998.txt.utf-8': 8, '/content/
dataset/pg1513.txt': 8, '/content/dataset/pg11.txt': 8, '/content/dataset/2160-0.txt': 2
5, '/content/dataset/345.txt.utf-8': 18, '/content/dataset/64317.txt.utf-8': 13, '/conte
nt/dataset/70817.txt.utf-8': 8, '/content/dataset/1080-0.txt': 8, '/content/dataset/pg26
00.txt': 29, '/content/dataset/70815.txt.utf-8': 10}, 'donations': {'/content/dataset/19
98.txt.utf-8': 15, '/content/dataset/pg1513.txt': 15, '/content/dataset/pg11.txt': 15,
'/content/dataset/2160-0.txt': 15, '/content/dataset/345.txt.utf-8': 15, '/content/data
set/64317.txt.utf-8': 15, '/content/dataset/70817.txt.utf-8': 15, '/content/dataset/1080
-0.txt': 15, '/content/dataset/pg2600.txt': 15, '/content/dataset/70815.txt.utf-8': 15},
'notifies': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/
content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.tx
t.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8':
1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset
/70815.txt.utf-8': 1}, 'email': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset
/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/co
ntent/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset
/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt':
2, '/content/dataset/70815.txt.utf-8': 2}, 'receipt': {'/content/dataset/1998.txt.utf-
8': 2, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 2, '/content/datas
et/2160-0.txt': 6, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-
8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/conten
t/dataset/pg2600.txt': 6, '/content/dataset/70815.txt.utf-8': 2}, 'replacement': {'/cont
ent/dataset/1998.txt.utf-8': 5, '/content/dataset/pg1513.txt': 5, '/content/dataset/pg1
1.txt': 5, '/content/dataset/2160-0.txt': 5, '/content/dataset/345.txt.utf-8': 5, '/cont
ent/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset
/1080-0.txt': 5, '/content/dataset/pg2600.txt': 5, '/content/dataset/70815.txt.utf-8':
5}, 'group': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1,
'/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/34
5.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.txt.utf-
```

[illegible]

```
content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.t
xt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/co
ntent/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/p
g2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}, 'computer': {'/content/dataset/19
98.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/co
ntent/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/643
17.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.txt':
2, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 2}, 'virus':
{'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/data
set/pg11.txt': 1, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 1,
'/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/conten
t/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.
utf-8': 1}, 'damage': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.tx
t': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 10, '/content/data
set/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.tx
t.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 3, '/conte
nt/dataset/70815.txt.utf-8': 3}, 'equipment': {'/content/dataset/1998.txt.utf-8': 3, '/c
ontent/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/dataset/2160-0.
txt': 3, '/content/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 3, '/c
ontent/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/content/dataset/
pg2600.txt': 11, '/content/dataset/70815.txt.utf-8': 4}, 'warranty': {'/content/dataset/
1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/
content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/6
4317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-0.tx
t': 2, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 2}, 'discla
imer': {'/content/dataset/1998.txt.utf-8': 3, '/content/dataset/pg1513.txt': 3, '/conten
t/dataset/pg11.txt': 3, '/content/dataset/2160-0.txt': 3, '/content/dataset/345.txt.utf-
8': 3, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/70817.txt.utf-8': 3, '/c
ontent/dataset/1080-0.txt': 3, '/content/dataset/pg2600.txt': 3, '/content/dataset/7081
5.txt.utf-8': 3}, 'damages': {'/content/dataset/1998.txt.utf-8': 4, '/content/dataset/pg
1513.txt': 4, '/content/dataset/pg11.txt': 4, '/content/dataset/2160-0.txt': 6, '/conten
t/dataset/345.txt.utf-8': 5, '/content/dataset/64317.txt.utf-8': 4, '/content/dataset/70
817.txt.utf-8': 4, '/content/dataset/1080-0.txt': 4, '/content/dataset/pg2600.txt': 4,
'/content/dataset/70815.txt.utf-8': 4}, 'liability': {'/content/dataset/1998.txt.utf-
8': 3, '/content/dataset/pg1513.txt': 3, '/content/dataset/pg11.txt': 3, '/content/datas
et/2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-
8': 3, '/content/dataset/70817.txt.utf-8': 3, '/content/dataset/1080-0.txt': 3, '/conten
t/dataset/pg2600.txt': 3, '/content/dataset/70815.txt.utf-8': 3}, 'costs': {'/content/da
taset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt':
2, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8': 2, '/content/dat
aset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2, '/content/dataset/1080-
0.txt': 2, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 2}, 'ex
penses': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/pg1513.txt': 2, '/cont
ent/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.ut
f-8': 2, '/content/dataset/64317.txt.utf-8': 2, '/content/dataset/70817.txt.utf-8': 2,
'/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 3, '/content/dataset/7
0815.txt.utf-8': 2}, 'legal': {'/content/dataset/1998.txt.utf-8': 2, '/content/dataset/p
g1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/2160-0.txt': 4, '/conte
nt/dataset/345.txt.utf-8': 6, '/content/dataset/64317.txt.utf-8': 3, '/content/dataset/7
0817.txt.utf-8': 2, '/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 8,
'/content/dataset/70815.txt.utf-8': 2}, 'remedies': {'/content/dataset/1998.txt.utf-8':
1, '/content/dataset/pg1513.txt': 2, '/content/dataset/pg11.txt': 2, '/content/dataset/
2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8':
1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/d
ataset/pg2600.txt': 3, '/content/dataset/70815.txt.utf-8': 1}, 'negligence': {'/content/
dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.tx
t': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/
dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/10
80-0.txt': 1, '/content/dataset/pg2600.txt': 3, '/content/dataset/70815.txt.utf-8': 1},
'distributor': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1,
'/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/34
5.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-
8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dat
aset/70815.txt.utf-8': 1}, 'liable': {'/content/dataset/1998.txt.utf-8': 1, '/content/da
taset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2,
'/content/dataset/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 3, '/content/d
```


[illegible]

[illegible]

[illegible]

[illegible]

[illegible]

```

2, '/content/dataset/pg2600.txt': 10, '/content/dataset/70815.txt.utf-8': 2}, 'checks':
{'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/data
set/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1,
'/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/conten
t/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.txt.
utf-8': 1}, 'card': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.tx
t': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 9, '/content/datas
et/345.txt.utf-8': 3, '/content/dataset/64317.txt.utf-8': 5, '/content/dataset/70817.tx
t.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 19, '/cont
ent/dataset/70815.txt.utf-8': 1}, 'professor': {'/content/dataset/1998.txt.utf-8': 1, '/
content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-
0.txt': 5, '/content/dataset/345.txt.utf-8': 175, '/content/dataset/64317.txt.utf-8': 1,
'/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/data
set/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 1}, 'michael': {'/content/datase
t/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1,
'/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.utf-8': 1, '/content/datase
t/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.t
xt': 1, '/content/dataset/pg2600.txt': 53, '/content/dataset/70815.txt.utf-8': 1}, 'har
t': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/d
ataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 4, '/content/dataset/345.txt.utf-8':
1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/cont
ent/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 1, '/content/dataset/70815.tx
t.utf-8': 1}, 'originator': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1
513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 1, '/conten
t/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70
817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 3,
'/content/dataset/70815.txt.utf-8': 1}, 'library': {'/content/dataset/1998.txt.utf-8':
1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/
2160-0.txt': 3, '/content/dataset/345.txt.utf-8': 9, '/content/dataset/64317.txt.utf-8':
11, '/content/dataset/70817.txt.utf-8': 5, '/content/dataset/1080-0.txt': 1, '/content/
dataset/pg2600.txt': 6, '/content/dataset/70815.txt.utf-8': 1}, 'network': {'/content/da
taset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt':
1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dat
aset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-
0.txt': 1, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 1}, 'vo
lunteer': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/con
tent/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 2, '/content/dataset/345.txt.u
tf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1,
'/content/dataset/1080-0.txt': 2, '/content/dataset/pg2600.txt': 3, '/content/dataset/7
0815.txt.utf-8': 1}, 'confirmed': {'/content/dataset/1998.txt.utf-8': 1, '/content/datas
et/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 11, '/
content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.utf-8': 3, '/content/data
set/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.tx
t': 20, '/content/dataset/70815.txt.utf-8': 1}, 'facility': {'/content/dataset/1998.txt.
utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/d
ataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 2, '/content/dataset/64317.txt.
utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/co
ncontent/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8': 2}, 'includes': {'/con
tent/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt': 1, '/content/dataset/pg1
1.txt': 1, '/content/dataset/2160-0.txt': 1, '/content/dataset/345.txt.utf-8': 1, '/cont
ent/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.utf-8': 1, '/content/datase
t/1080-0.txt': 1, '/content/dataset/pg2600.txt': 2, '/content/dataset/70815.txt.utf-8':
1}, 'subscribe': {'/content/dataset/1998.txt.utf-8': 1, '/content/dataset/pg1513.txt':
1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.txt': 4, '/content/dataset/
345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/content/dataset/70817.txt.ut
f-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg2600.txt': 2, '/content/d
ataset/70815.txt.utf-8': 1}, 'newsletter': {'/content/dataset/1998.txt.utf-8': 1, '/cont
ent/dataset/pg1513.txt': 1, '/content/dataset/pg11.txt': 1, '/content/dataset/2160-0.tx
t': 1, '/content/dataset/345.txt.utf-8': 1, '/content/dataset/64317.txt.utf-8': 1, '/con
tent/dataset/70817.txt.utf-8': 1, '/content/dataset/1080-0.txt': 1, '/content/dataset/pg
2600.txt': 1, '/content/dataset/70815.txt.utf-8': 1}}

```

Words are any sequence of alphanumeric characters (i.e., letters and numbers) separated by space which has particular meaning. So, here I have considered a website address as a word. As per the above code I have used `SparkContext.getOrCreate()` to either get the current Spark context or to create one if it does not exist.

Then `sc.textFile(file)` is written to load the file as an RDD of strings. It applies a `flatMap` operation on the RDD, where it applies a regular expression to each line to extract words, converting all words to lower case. `intersection()` function returns an RDD that contains the intersection of elements in the source RDD and the other RDD, `filter()` operation returns a new RDD containing only the elements that satisfy a predicate, `countByKey()` function is then called to count the number of occurrences of each unique value in the RDD.

Question: 03

Perform an exploratory data analysis on the `primary_data.csv` of the Mushroom dataset [Secondary Mushroom Dataset](#) using pandas and answer the following questions:

- How many mushrooms are poisonous?
- Which family contains most poisonous mushrooms?
- Do all poisonous mushrooms have a ring?
- Which seasons do poisonous mushrooms grow typically?

3.a) How many mushrooms are poisonous?

```
In [ ]: import pandas as pd
df= pd.read_csv('/content/dataset/primary_data_edited.csv',delimiter=";")
df.head(3)
```

```
Out[ ]:
```

	family	name	class	cap-diameter	cap-shape	Cap-surface	cap-color	does-bruise-or-bleed	gill-attachment	gill-spacing	gill-color	stem-height	stem-width	stem-root	stem-surface	stem-color
0	Amanita Family	Fly Agaric	p	[10, 20]	[x, f]	[g, h]	[e, o]	[f]	[e]	NaN	...	[s]	[y]			
1	Amanita Family	Panther Cap	p	[5, 10]	[p, x]	[g]	[n]	[f]	[e]	NaN	...	NaN	[y]			
2	Amanita Family	False Panther Cap	p	[10, 15]	[x, f]	NaN	[g, n]	[f]	[e]	NaN	...	NaN	NaN			

3 rows × 23 columns

```
In [ ]: df.columns
```

```
Out[ ]: Index(['family', 'name', 'class', 'cap-diameter', 'cap-shape', 'Cap-surface',
        'cap-color', 'does-bruise-or-bleed', 'gill-attachment', 'gill-spacing',
        'gill-color', 'stem-height', 'stem-width', 'stem-root', 'stem-surface',
        'stem-color', 'veil-type', 'veil-color', 'has-ring', 'ring-type',
        'Spore-print-color', 'habitat', 'season'],
        dtype='object')
```

```
In [ ]: print(f"Number of mushrooms are poisonous: {len(df[df['class'] == 'p'])}")
```

Number of mushrooms are poisonous: 96

3.b) Which family contains most poisonous mushrooms?

```
In [ ]: print(f"The family name contains most poisonous mushrooms is: {df[df['class'] == 'p']['fa']
The family name contains most poisonous mushrooms is: Tricholoma Family
```

3.c) Do all poisonous mushrooms have a ring?

```
In [ ]: df[df['class'] == 'p'][df['has-ring'] == 'f']
```

<ipython-input-67-2659131bd551>:1: UserWarning: Boolean Series key will be reindexed to match DataFrame index.

```
df[df['class'] == 'p'][df['has-ring'] == 'f']
```

Out[]:

	family	name	class	cap- diameter	cap- shape	Cap- surface	cap- color	does- bruise- or- bleed	gill- attachment	gill- spacing	...	stem- root	ster- surfa
17	Tricholoma Family	Club- footed Funnel Cap	p	[4, 8]	[x, f]	NaN	[g, n]	[f]	[d]	[c]	...	[b]	Na
20	Tricholoma Family	Ivory Clitocybe	p	[2, 4]	[x, f, s]	[s]	[w, g]	[f]	[d]	[c]	...	NaN	Na
21	Tricholoma Family	Wood Woolly- foot	p	[3, 6]	[f]	[l]	[n, e, y]	[f]	[x]	[d]	...	NaN	Na
22	Tricholoma Family	Russet Tough Shank	p	[2, 5]	[f]	[d]	[n, b]	[f]	[x]	[c]	...	[b]	[i,
23	Tricholoma Family	Clustered Tough Shank	p	[2, 5]	[x, f]	NaN	[n, w]	[f]	[x]	[c]	...	NaN	Na
...
164	Bracket Fungi	Silver Leaf Fungus	p	[2, 4]	[o]	[l]	[g]	[f]	[f]	[f]	...	[f]	
166	Bracket Fungi	Fuzzy Polypore	p	[8, 30]	[f, x]	[g, s, d]	[o, e, n, k]	[f]	[p]	NaN	...	NaN	Na
169	Bracket Fungi	Elfin's Saddle	p	[2, 4]	[o]	NaN	[g, k]	[f]	[f]	[f]	...	NaN	[
170	Bracket Fungi	Turban Fungus	p	[5, 15]	[o]	NaN	[n, e]	[f]	[f]	[f]	...	NaN	[
172	Jelly Discs Family	Jelly Babies	p	[1, 1.5]	[x, f, s]	[s]	[y]	[f]	[f]	[f]	...	NaN	Na

70 rows × 23 columns

No, all poisonous mushrooms don't have a ring. 70 mushrooms don't have a ring.

3.d) Which seasons do poisonous mushrooms grow typically?

```
In [ ]: season_name=df[df['class'] == 'p']['season'].value_counts().idxmax()
new_season=season_name.replace("s", "spring")
new_season=new_season.replace("u", "summer")
new_season=new_season.replace("a", "autumn")
new_season=new_season.replace("w", "winter")
print(f"The poisonous mushrooms grow typically in: {new_season} seasons")
```

The poisonous mushrooms grow typically in: [summer, autumn] seasons

So, I got The poisonous mushrooms grow typically in: [summer, autumn] seasons

Hint:

To download files on colab or on any notebook environment without a direct terminal access, you can use **wget**

For Eg:

```
In [ ]: # First make sure the directory you are supposed to download to, is available and access
!mkdir -p dataset
!wget -P dataset https://www.gutenberg.org/cache/epub/1513/pg1513.txt # Here the -P flag

--2023-05-30 18:48:44--  https://www.gutenberg.org/cache/epub/1513/pg1513.txt
Resolving www.gutenberg.org (www.gutenberg.org)... 152.19.134.47, 2610:28:3090:3000:0:ba
d:cafe:47
Connecting to www.gutenberg.org (www.gutenberg.org)|152.19.134.47|:443... connected.
HTTP request sent, awaiting response... 200 OK
Length: 169260 (165K) [text/plain]
Saving to: 'dataset/pg1513.txt'

pg1513.txt          100%[=====>] 165.29K   990KB/s   in 0.2s

2023-05-30 18:48:44 (990 KB/s) - 'dataset/pg1513.txt' saved [169260/169260]
```

Now do the same for all dataset links you encounter.