

MapReduce-WordCount (1)

July 16, 2025

```
[1]: from functools import reduce
import string
```

```
[2]: translator = str.maketrans('', '', string.punctuation)
```

```
[9]: 'test!*'.translate(translator)
```

```
[9]: 'test'
```

```
[3]: words_lst = open('shakespeare.txt')\
        .read()\
        .split()
```

```
[5]: words_lst
```

```
[5]: ['This',
      'is',
      'the',
      '100th',
      'Etext',
      'file',
      'presented',
      'by',
      'Project',
      'Gutenberg,',
      'and',
      'is',
      'presented',
      'in',
      'cooperation',
      'with',
      'World',
      'Library,',
      'Inc.,',
      'from',
      'their',
      'Library',
```

'of',
'the',
'Future',
'and',
'Shakespeare',
'CDROMS.',
'Project',
'Gutenberg',
'often',
'releases',
'Etexts',
'that',
'are',
'NOT',
'placed',
'in',
'the',
'Public',
'Domain!!',
'Shakespeare',
'*This',
'Etext',
'has',
'certain',
'copyright',
'implications',
'you',
'should',
'read!*',
'<<THIS',
'ELECTRONIC',
'VERSION',
'OF',
'THE',
'COMPLETE',
'WORKS',
'OF',
'WILLIAM',
'SHAKESPEARE',
'IS',
'COPYRIGHT',
'1990-1993',
'BY',
'WORLD',
'LIBRARY,',
'INC.',
'AND',

'IS',
'PROVIDED',
'BY',
'PROJECT',
'GUTENBERG',
'ETEXT',
'OF',
'ILLINOIS',
'BENEDICTINE',
'COLLEGE',
'WITH',
'PERMISSION.',
'ELECTRONIC',
'AND',
'MACHINE',
'READABLE',
'COPIES',
'MAY',
'BE',
'DISTRIBUTED',
'SO',
'LONG',
'AS',
'SUCH',
'COPIES',
'(1)',
'ARE',
'FOR',
'YOUR',
'OR',
'OTHERS',
'PERSONAL',
'USE',
'ONLY',
'AND',
'(2)',
'ARE',
'NOT',
'DISTRIBUTED',
'OR',
'USED',
'COMMERCIALLY.',
'PROHIBITED',
'COMMERCIAL',
'DISTRIBUTION',
'INCLUDES',
'BY',

'ANY',
'SERVICE',
'THAT',
'CHARGES',
'FOR',
'DOWNLOAD',
'TIME',
'OR',
'FOR',
'MEMBERSHIP.>>',
'*Project',
'Gutenberg',
'is',
'proud',
'to',
'cooperate',
'with',
'The',
'World',
'Library*',
'in',
'the',
'presentation',
'of',
'The',
'Complete',
'Works',
'of',
'William',
'Shakespeare',
'for',
'your',
'reading',
'for',
'education',
'and',
'entertainment.',
'HOWEVER,',
'THIS',
'IS',
'NEITHER',
'SHAREWARE',
'NOR',
'PUBLIC',
'DOMAIN.',
'.',
'AND',

'UNDER',
'THE',
'LIBRARY',
'OF',
'THE',
'FUTURE',
'CONDITIONS',
'OF',
'THIS',
'PRESENTATION.',
'.',
'NO',
'CHARGES',
'MAY',
'BE',
'MADE',
'FOR',
'*ANY*',
'ACCESS',
'TO',
'THIS',
'MATERIAL.',
'YOU',
'ARE',
'ENCOURAGED!!!',
'TO',
'GIVE',
'IT',
'AWAY',
'TO',
'ANYONE',
'YOU',
'LIKE',
'BUT',
'NO',
'CHARGES',
'ARE',
'ALLOWED!!!',
'**Welcome',
'To',
'The',
'World',
'of',
'Free',
'Plain',
'Vanilla',
'Electronic',

'Texts**',
'**Etexts',
'Readable',
'By',
'Both',
'Humans',
'and',
'By',
'Computers,',
'Since',
'1971**',
'*These',
'Etexts',
'Prepared',
'By',
'Hundreds',
'of',
'Volunteers',
'and',
'Donations*',
'Information',
'on',
'contacting',
'Project',
'Gutenberg',
'to',
'get',
'Etexts,',
'and',
'further',
'information',
'is',
'included',
'below.',
'We',
'need',
'your',
'donations.',
'The',
'Complete',
'Works',
'of',
'William',
'Shakespeare',
'January,',
'1994',
' [Etext',

'#100]',
'The',
'Library',
'of',
'the',
'Future',
'Complete',
'Works',
'of',
'William',
'Shakespeare',
'Library',
'of',
'the',
'Future',
'is',
'a',
'TradeMark',
'(TM)',
'of',
'World',
'Library',
'Inc.',
'*****This',
'file',
'should',
'be',
'named',
'shaks12.txt',
'or',
'shaks12.zip*****',
'Corrected',
'EDITIONS',
'of',
'our',
'etexts',
'get',
'a',
'new',
'NUMBER',
'shaks13.txt',
'VERSIONS',
'based',
'on',
'separate',
'sources',
'get',

'new',
'LETTER',
'shaks10a.txt',
'If',
'you',
'would',
'like',
'further',
'information',
'about',
'World',
'Library',
'Inc.',
'Please',
'call',
'them',
'at',
'1-800-443-0238',
'or',
'email',
'julianc@netcom.com',
'Please',
'give',
'them',
'our',
'thanks',
'for',
'their',
'Shakespeare',
'cooperation!',
'The',
'official',
'release',
'date',
'of',
'all',
'Project',
'Gutenberg',
'Etexts',
'is',
'at',
'Midnight',
'Central',
'Time',
'of',
'the',
'last',

'day',
'of',
'the',
'stated',
'month.',
'A',
'preliminary',
'version',
'may',
'often',
'be',
'posted',
'for',
'suggestion,',
'comment',
'and',
'editing',
'by',
'those',
'who',
'wish',
'to',
'do',
'so.',
'To',
'be',
'sure',
'you',
'have',
'an',
'up',
'to',
'date',
'first',
'edition',
'[xxxxx10x.xxx]',
'please',
'check',
'file',
'sizes',
'in',
'the',
'first',
'week',
'of',
'the',
'next',

'month.',
'Since',
'our',
'ftp',
'program',
'has',
'a',
'bug',
'in',
'it',
'that',
'scrambles',
'the',
'date',
'[tried',
'to',
'fix',
'and',
'failed]',
'a',
'look',
'at',
'the',
'file',
'size',
'will',
'have',
'to',
'do',
'but',
'we',
'will',
'try',
'to',
'see',
'a',
'new',
'copy',
'has',
'at',
'least',
'one',
'byte',
'more',
'or',
'less.',
'Information',

'about',
'Project',
'Gutenberg',
'(one',
'page)',
'We',
'produce',
'about',
'two',
'million',
'dollars',
'for',
'each',
'hour',
'we',
'work.',
'The',
'fifty',
'hours',
'is',
'one',
'conservative',
'estimate',
'for',
'how',
'long',
'it',
'we',
'take',
'to',
'get',
'any',
'etext',
'selected,',
'entered,',
'proofread,',
'edited,',
'copyright',
'searched',
'and',
'analyzed,',
'the',
'copyright',
'letters',
'written,',
'etc.',
'This',

'projected',
'audience',
'is',
'one',
'hundred',
'million',
'readers.',
'If',
'our',
'value',
'per',
'text',
'is',
'nominally',
'estimated',
'at',
'one',
'dollar,',
'then',
'we',
'produce',
'2',
'million',
'dollars',
'per',
'hour',
'this',
'year',
'we,',
'will',
'have',
'to',
'do',
'four',
'text',
'files',
'per',
'month:',
'thus',
'upping',
'our',
'productivity',
'from',
'one',
'million.',
'The',
'Goal',

'of',
'Project',
'Gutenberg',
'is',
'to',
'Give',
'Away',
'One',
'Trillion',
'Etext',
'Files',
'by',
'the',
'December',
'31,',
'2001.',
'[10,000',
'x',
'100,000,000=Trillion] ',
'This',
'is',
'ten',
'thousand',
'titles',
'each',
'to',
'one',
'hundred',
'million',
'readers,',
'which',
'is',
'10%',
'of',
'the',
'expected',
'number',
'of',
'computer',
'users',
'by',
'the',
'end',
'of',
'the',
'year',
'2001.',

'We',
'need',
'your',
'donations',
'more',
'than',
'ever!',
'All',
'donations',
'should',
'be',
'made',
'to',
'"Project',
'Gutenberg/IBC"',
'and',
'are',
'tax',
'deductible',
'to',
'the',
'extent',
'allowable',
'by',
'law',
'("IBC"',
'is',
'Illinois',
'Benedictine',
'College).',
'(Subscriptions',
'to',
'our',
'paper',
'newsletter',
'go',
'to',
'IBC',
'too)',
'For',
'these',
'and',
'other',
'matters',
'please',
'mail',
'to:',

'Project',
'Gutenberg',
'P.',
'O.',
'Box',
'2782',
'Champaign,',
'IL',
'61825',
'When',
'all',
'other',
'email',
'fails',
'try',
'our',
'Michael',
'S.',
'Hart,',
'Executive',
'Director:',
'hart@vmd.cso.uiuc.edu',
'(internet)',
'hart@uiucvmd',
'(bitnet)',
'We',
'would',
'prefer',
'to',
'send',
'you',
'this',
'information',
'by',
'email',
'(Internet,',
'Bitnet,',
'Compuserve,',
'ATTMAIL',
'or',
'MCImail).',
'*****',
'If',
'you',
'have',
'an',
'FTP',

'program',
'(or',
'emulator)',',
'please',
'FTP',
'directly',
'to',
'the',
'Project',
'Gutenberg',
'archives:',
'[Mac',
'users,',
'do',
'NOT',
'point',
'and',
'click.',
'.',
'type]',
'ftp',
'mrcnext.cso.uiuc.edu',
'login:',
'anonymous',
'password:',
'your@login',
'cd',
'etext/etext91',
'or',
'cd',
'etext92',
'or',
'cd',
'etext93',
'[for',
'new',
'books]',
'[now',
'also',
'in',
'cd',
'etext/etext93]',
'or',
'cd',
'etext/articles',
'[get',
'suggest',

'gut',
'for',
'more',
'information]',
'dir',
'[to',
'see',
'files]',
'get',
'or',
'mget',
'[to',
'get',
'files.',
'.',
'set',
'bin',
'for',
'zip',
'files]',
'GET',
'OINDEX.GUT',
'for',
'a',
'list',
'of',
'books',
'and',
'GET',
'NEW',
'GUT',
'for',
'general',
'information',
'and',
'MGET',
'GUT*',
'for',
'newsletters.',
'**Information',
'prepared',
'by',
'the',
'Project',
'Gutenberg',
'legal',
'advisor**',

'*****',
'SMALL',
'PRINT! ',
'for',
'COMPLETE',
'SHAKESPEARE',
'*****',
'THIS',
'ELECTRONIC',
'VERSION',
'OF',
'THE',
'COMPLETE',
'WORKS',
'OF',
'WILLIAM',
'SHAKESPEARE',
'IS',
'COPYRIGHT',
'1990-1993',
'BY',
'WORLD',
'LIBRARY, ',
'INC., ',
'AND',
'IS',
'PROVIDED',
'BY',
'PROJECT',
'GUTENBERG',
'ETEXT',
'OF',
'ILLINOIS',
'BENEDICTINE',
'COLLEGE',
'WITH',
'PERMISSION.',
'Since',
'unlike',
'many',
'other',
'Project',
'Gutenberg-tm',
'etexts,',
'this',
'etext',
'is',

'copyright',
'protected',
'and',
'since',
'the',
'materials',
'and',
'methods',
'you',
'use',
'will',
'effect',
'the',
"Project's",
'reputation',
'your',
'right',
'to',
'copy',
'and',
'distribute',
'it',
'is',
'limited',
'by',
'the',
'copyright',
'and',
'other',
'laws',
'and',
'by',
'the',
'conditions',
'of',
'this',
"Small",
'Print!',
'statement.',
'1.',
'LICENSE',
'A)',
'YOU',
'MAY',
'(AND',
'ARE',
'ENCOURAGED)',

'TO',
'DISTRIBUTE',
'ELECTRONIC',
'AND',
'MACHINE',
'READABLE',
'COPIES',
'OF',
'THIS',
'ETEXT',
'SO',
'LONG',
'AS',
'SUCH',
'COPIES',
'(1)',
'ARE',
'FOR',
'YOUR',
'OR',
'OTHERS',
'PERSONAL',
'USE',
'ONLY',
'AND',
'(2)',
'ARE',
'NOT',
'DISTRIBUTED',
'OR',
'USED',
'COMMERCIALLY.',
'PROHIBITED',
'COMMERCIAL',
'DISTRIBUTION',
'INCLUDES',
'BY',
'ANY',
'SERVICE',
'THAT',
'CHARGES',
'FOR',
'DOWNLOAD',
'TIME',
'OR',
'FOR',
'MEMBERSHIP.',

'B)',
'This',
'license',
'is',
'subject',
'to',
'the',
'conditions',
'that',
'you',
'honor',
'the',
'refund',
'and',
'replacement',
'provisions',
'of',
'this',
"small",
'print!',
'statement;',
'and',
'that',
'you',
'distribute',
'exact',
'copies',
'of',
'this',
'etext,',
'including',
'this',
'Small',
'Print',
'statement.',
'Such',
'copies',
'can',
'be',
'compressed',
'or',
'any',
'proprietary',
'form',
'(including',
'any',
'form',

```

'resulting',
'from',
'word',
'processing',
'or',
'hypertext',
'software)',
'so',
'long',
'as',
'*EITHER*',
'(1)',
'The',
'etext,',
'when',
'displayed,',
'is',
'clearly',
'readable,',
'and',
'does',
'*not*',
'contain',
'characters',
'other',
'than',
'those',
'intended',
'by',
'the',
'author',
'of',
'the',
'work,',
'although',
'tilde',
'(~)',
'asterisk',
...]
```

```
[4]: unified_lst = list(map(lambda word: word.translate(translator)\
                             .lower()\
                             , words_lst))
```

```

#def cleanse(word):
#    word = word.translate(translator)
```

```
# word = word.lower()
# return word
```

```
[7]: unified_lst
```

```
[7]: ['this',
      'is',
      'the',
      '100th',
      'etext',
      'file',
      'presented',
      'by',
      'project',
      'guttenberg',
      'and',
      'is',
      'presented',
      'in',
      'cooperation',
      'with',
      'world',
      'library',
      'inc',
      'from',
      'their',
      'library',
      'of',
      'the',
      'future',
      'and',
      'shakespeare',
      'cdroms',
      'project',
      'guttenberg',
      'often',
      'releases',
      'etexts',
      'that',
      'are',
      'not',
      'placed',
      'in',
      'the',
      'public',
      'domain',
      'shakespeare',
```

'this',
'etext',
'has',
'certain',
'copyright',
'implications',
'you',
'should',
'read',
'this',
'electronic',
'version',
'of',
'the',
'complete',
'works',
'of',
'william',
'shakespeare',
'is',
'copyright',
'19901993',
'by',
'world',
'library',
'inc',
'and',
'is',
'provided',
'by',
'project',
'gutenberg',
'etext',
'of',
'illinois',
'benedictine',
'college',
'with',
'permission',
'electronic',
'and',
'machine',
'readable',
'copies',
'may',
'be',
'distributed',

'so',
'long',
'as',
'such',
'copies',
'1',
'are',
'for',
'your',
'or',
'others',
'personal',
'use',
'only',
'and',
'2',
'are',
'not',
'distributed',
'or',
'used',
'commercially',
'prohibited',
'commercial',
'distribution',
'includes',
'by',
'any',
'service',
'that',
'charges',
'for',
'download',
'time',
'or',
'for',
'membership',
'project',
'gutenberg',
'is',
'proud',
'to',
'cooperate',
'with',
'the',
'world',
'library',

'in',
'the',
'presentation',
'of',
'the',
'complete',
'works',
'of',
'william',
'shakespeare',
'for',
'your',
'reading',
'for',
'education',
'and',
'entertainment',
'however',
'this',
'is',
'neither',
'shareware',
'nor',
'public',
'domain',
'',
'and',
'under',
'the',
'library',
'of',
'the',
'future',
'conditions',
'of',
'this',
'presentation',
'',
'no',
'charges',
'may',
'be',
'made',
'for',
'any',
'access',
'to',

'this',
'material',
'you',
'are',
'encouraged',
'to',
'give',
'it',
'away',
'to',
'anyone',
'you',
'like',
'but',
'no',
'charges',
'are',
'allowed',
'welcome',
'to',
'the',
'world',
'of',
'free',
'plain',
'vanilla',
'electronic',
'texts',
'etexts',
'readable',
'by',
'both',
'humans',
'and',
'by',
'computers',
'since',
'1971',
'these',
'etexts',
'prepared',
'by',
'hundreds',
'of',
'volunteers',
'and',
'donations',

'information',
'on',
'contacting',
'project',
'gutenberg',
'to',
'get',
'etexts',
'and',
'further',
'information',
'is',
'included',
'below',
'we',
'need',
'your',
'donations',
'the',
'complete',
'works',
'of',
'william',
'shakespeare',
'january',
'1994',
'etext',
'100',
'the',
'library',
'of',
'the',
'future',
'complete',
'works',
'of',
'william',
'shakespeare',
'library',
'of',
'the',
'future',
'is',
'a',
'trademark',
'tm',
'of',

'world',
'library',
'inc',
'this',
'file',
'should',
'be',
'named',
'shaks12txt',
'or',
'shaks12zip',
'corrected',
'editions',
'of',
'our',
'etexts',
'get',
'a',
'new',
'number',
'shaks13txt',
'versions',
'based',
'on',
'separate',
'sources',
'get',
'new',
'letter',
'shaks10atxt',
'if',
'you',
'would',
'like',
'further',
'information',
'about',
'world',
'library',
'inc',
'please',
'call',
'them',
'at',
'18004430238',
'or',
'email',

'juliancnetcomcom',
'please',
'give',
'them',
'our',
'thanks',
'for',
'their',
'shakespeare',
'cooperation',
'the',
'official',
'release',
'date',
'of',
'all',
'project',
'gutenberg',
'etexts',
'is',
'at',
'midnight',
'central',
'time',
'of',
'the',
'last',
'day',
'of',
'the',
'stated',
'month',
'a',
'preliminary',
'version',
'may',
'often',
'be',
'posted',
'for',
'suggestion',
'comment',
'and',
'editing',
'by',
'those',
'who',

'wish',
'to',
'do',
'so',
'to',
'be',
'sure',
'you',
'have',
'an',
'up',
'to',
'date',
'first',
'edition',
'xxxxx10xxxx',
'please',
'check',
'file',
'sizes',
'in',
'the',
'first',
'week',
'of',
'the',
'next',
'month',
'since',
'our',
'ftp',
'program',
'has',
'a',
'bug',
'in',
'it',
'that',
'scrambles',
'the',
'date',
'tried',
'to',
'fix',
'and',
'failed',
'a',

'look',
'at',
'the',
'file',
'size',
'will',
'have',
'to',
'do',
'but',
'we',
'will',
'try',
'to',
'see',
'a',
'new',
'copy',
'has',
'at',
'least',
'one',
'byte',
'more',
'or',
'less',
'information',
'about',
'project',
'gutenberg',
'one',
'page',
'we',
'produce',
'about',
'two',
'million',
'dollars',
'for',
'each',
'hour',
'we',
'work',
'the',
'fifty',
'hours',
'is',

'one',
'conservative',
'estimate',
'for',
'how',
'long',
'it',
'we',
'take',
'to',
'get',
'any',
'etext',
'selected',
'entered',
'proofread',
'edited',
'copyright',
'searched',
'and',
'analyzed',
'the',
'copyright',
'letters',
'written',
'etc',
'this',
'projected',
'audience',
'is',
'one',
'hundred',
'million',
'readers',
'if',
'our',
'value',
'per',
'text',
'is',
'nominally',
'estimated',
'at',
'one',
'dollar',
'then',
'we',

'produce',
'2',
'million',
'dollars',
'per',
'hour',
'this',
'year',
'we',
'will',
'have',
'to',
'do',
'four',
'text',
'files',
'per',
'month',
'thus',
'upping',
'our',
'productivity',
'from',
'one',
'million',
'the',
'goal',
'of',
'project',
'gutenberg',
'is',
'to',
'give',
'away',
'one',
'trillion',
'etext',
'files',
'by',
'the',
'december',
'31',
'2001',
'10000',
'x',
'100000000trillion',
'this',

'is',
'ten',
'thousand',
'titles',
'each',
'to',
'one',
'hundred',
'million',
'readers',
'which',
'is',
'10',
'of',
'the',
'expected',
'number',
'of',
'computer',
'users',
'by',
'the',
'end',
'of',
'the',
'year',
'2001',
'we',
'need',
'your',
'donations',
'more',
'than',
'ever',
'all',
'donations',
'should',
'be',
'made',
'to',
'project',
'gutenbergibc',
'and',
'are',
'tax',
'deductible',
'to',

'the',
'extent',
'allowable',
'by',
'law',
'ibc',
'is',
'illinois',
'benedictine',
'college',
'subscriptions',
'to',
'our',
'paper',
'newsletter',
'go',
'to',
'ibc',
'too',
'for',
'these',
'and',
'other',
'matters',
'please',
'mail',
'to',
'project',
'gutenberg',
'p',
'o',
'box',
'2782',
'champaign',
'il',
'61825',
'when',
'all',
'other',
'email',
'fails',
'try',
'our',
'michael',
's',
'hart',
'executive',

'director',
'hartvmdcsouiucedu',
'internet',
'hartuiucvmd',
'bitnet',
'we',
'would',
'prefer',
'to',
'send',
'you',
'this',
'information',
'by',
'email',
'internet',
'bitnet',
'compuserve',
'attmail',
'or',
'mcimail',
'',
'if',
'you',
'have',
'an',
'ftp',
'program',
'or',
'emulator',
'please',
'ftp',
'directly',
'to',
'the',
'project',
'gutenberg',
'archives',
'mac',
'users',
'do',
'not',
'point',
'and',
'click',
'',
'type',

'ftp',
'mrcnextcsouiucedu',
'login',
'anonymous',
'password',
'yourlogin',
'cd',
'etextetext91',
'or',
'cd',
'etext92',
'or',
'cd',
'etext93',
'for',
'new',
'books',
'now',
'also',
'in',
'cd',
'etextetext93',
'or',
'cd',
'etextarticles',
'get',
'suggest',
'gut',
'for',
'more',
'information',
'dir',
'to',
'see',
'files',
'get',
'or',
'mget',
'to',
'get',
'files',
'',
'set',
'bin',
'for',
'zip',
'files',

'get',
'Oindexgut',
'for',
'a',
'list',
'of',
'books',
'and',
'get',
'new',
'gut',
'for',
'general',
'information',
'and',
'mget',
'gut',
'for',
'newsletters',
'information',
'prepared',
'by',
'the',
'project',
'gutenberg',
'legal',
'advisor',
'',
'small',
'print',
'for',
'complete',
'shakespeare',
'',
'this',
'electronic',
'version',
'of',
'the',
'complete',
'works',
'of',
'william',
'shakespeare',
'is',
'copyright',
'19901993',

'by',
'world',
'library',
'inc',
'and',
'is',
'provided',
'by',
'project',
'gutenberg',
'etext',
'of',
'illinois',
'benedictine',
'college',
'with',
'permission',
'since',
'unlike',
'many',
'other',
'project',
'gutenbergtm',
'etexts',
'this',
'etext',
'is',
'copyright',
'protected',
'and',
'since',
'the',
'materials',
'and',
'methods',
'you',
'use',
'will',
'effect',
'the',
'projects',
'reputation',
'your',
'right',
'to',
'copy',
'and',

'distribute',
'it',
'is',
'limited',
'by',
'the',
'copyright',
'and',
'other',
'laws',
'and',
'by',
'the',
'conditions',
'of',
'this',
'small',
'print',
'statement',
'1',
'license',
'a',
'you',
'may',
'and',
'are',
'encouraged',
'to',
'distribute',
'electronic',
'and',
'machine',
'readable',
'copies',
'of',
'this',
'etext',
'so',
'long',
'as',
'such',
'copies',
'1',
'are',
'for',
'your',
'or',

'others',
'personal',
'use',
'only',
'and',
'2',
'are',
'not',
'distributed',
'or',
'used',
'commercially',
'prohibited',
'commercial',
'distribution',
'includes',
'by',
'any',
'service',
'that',
'charges',
'for',
'download',
'time',
'or',
'for',
'membership',
'b',
'this',
'license',
'is',
'subject',
'to',
'the',
'conditions',
'that',
'you',
'honor',
'the',
'refund',
'and',
'replacement',
'provisions',
'of',
'this',
'small',
'print',

'statement',
'and',
'that',
'you',
'distribute',
'exact',
'copies',
'of',
'this',
'etext',
'including',
'this',
'small',
'print',
'statement',
'such',
'copies',
'can',
'be',
'compressed',
'or',
'any',
'proprietary',
'form',
'including',
'any',
'form',
'resulting',
'from',
'word',
'processing',
'or',
'hypertext',
'software',
'so',
'long',
'as',
'either',
'1',
'the',
'etext',
'when',
'displayed',
'is',
'clearly',
'readable',
'and',

```
'does',  
'not',  
'contain',  
'characters',  
'other',  
'than',  
'those',  
'intended',  
'by',  
'the',  
'author',  
'of',  
'the',  
'work',  
'although',  
'tilde',  
'',  
'asterisk',  
...]
```

```
[5]: map_lst = list(map(lambda word: dict([[word, 1]]), unified_lst))
```

```
[6]: map_lst
```

```
[6]: [{'this': 1},  
      {'is': 1},  
      {'the': 1},  
      {'100th': 1},  
      {'etext': 1},  
      {'file': 1},  
      {'presented': 1},  
      {'by': 1},  
      {'project': 1},  
      {'gutenberg': 1},  
      {'and': 1},  
      {'is': 1},  
      {'presented': 1},  
      {'in': 1},  
      {'cooperation': 1},  
      {'with': 1},  
      {'world': 1},  
      {'library': 1},  
      {'inc': 1},  
      {'from': 1},  
      {'their': 1},  
      {'library': 1},  
      {'of': 1},
```

{'the': 1},
{'future': 1},
{'and': 1},
{'shakespeare': 1},
{'cdroms': 1},
{'project': 1},
{'gutenberg': 1},
{'often': 1},
{'releases': 1},
{'etexts': 1},
{'that': 1},
{'are': 1},
{'not': 1},
{'placed': 1},
{'in': 1},
{'the': 1},
{'public': 1},
{'domain': 1},
{'shakespeare': 1},
{'this': 1},
{'etext': 1},
{'has': 1},
{'certain': 1},
{'copyright': 1},
{'implications': 1},
{'you': 1},
{'should': 1},
{'read': 1},
{'this': 1},
{'electronic': 1},
{'version': 1},
{'of': 1},
{'the': 1},
{'complete': 1},
{'works': 1},
{'of': 1},
{'william': 1},
{'shakespeare': 1},
{'is': 1},
{'copyright': 1},
{'19901993': 1},
{'by': 1},
{'world': 1},
{'library': 1},
{'inc': 1},
{'and': 1},
{'is': 1},

{'provided': 1},
{'by': 1},
{'project': 1},
{'gutenberg': 1},
{'etext': 1},
{'of': 1},
{'illinois': 1},
{'benedictine': 1},
{'college': 1},
{'with': 1},
{'permission': 1},
{'electronic': 1},
{'and': 1},
{'machine': 1},
{'readable': 1},
{'copies': 1},
{'may': 1},
{'be': 1},
{'distributed': 1},
{'so': 1},
{'long': 1},
{'as': 1},
{'such': 1},
{'copies': 1},
{'1': 1},
{'are': 1},
{'for': 1},
{'your': 1},
{'or': 1},
{'others': 1},
{'personal': 1},
{'use': 1},
{'only': 1},
{'and': 1},
{'2': 1},
{'are': 1},
{'not': 1},
{'distributed': 1},
{'or': 1},
{'used': 1},
{'commercially': 1},
{'prohibited': 1},
{'commercial': 1},
{'distribution': 1},
{'includes': 1},
{'by': 1},
{'any': 1},

{'service': 1},
{'that': 1},
{'charges': 1},
{'for': 1},
{'download': 1},
{'time': 1},
{'or': 1},
{'for': 1},
{'membership': 1},
{'project': 1},
{'gutenberg': 1},
{'is': 1},
{'proud': 1},
{'to': 1},
{'cooperate': 1},
{'with': 1},
{'the': 1},
{'world': 1},
{'library': 1},
{'in': 1},
{'the': 1},
{'presentation': 1},
{'of': 1},
{'the': 1},
{'complete': 1},
{'works': 1},
{'of': 1},
{'william': 1},
{'shakespeare': 1},
{'for': 1},
{'your': 1},
{'reading': 1},
{'for': 1},
{'education': 1},
{'and': 1},
{'entertainment': 1},
{'however': 1},
{'this': 1},
{'is': 1},
{'neither': 1},
{'shareware': 1},
{'nor': 1},
{'public': 1},
{'domain': 1},
{'': 1},
{'and': 1},
{'under': 1},

{ 'the': 1 },
{ 'library': 1 },
{ 'of': 1 },
{ 'the': 1 },
{ 'future': 1 },
{ 'conditions': 1 },
{ 'of': 1 },
{ 'this': 1 },
{ 'presentation': 1 },
{ '': 1 },
{ 'no': 1 },
{ 'charges': 1 },
{ 'may': 1 },
{ 'be': 1 },
{ 'made': 1 },
{ 'for': 1 },
{ 'any': 1 },
{ 'access': 1 },
{ 'to': 1 },
{ 'this': 1 },
{ 'material': 1 },
{ 'you': 1 },
{ 'are': 1 },
{ 'encouraged': 1 },
{ 'to': 1 },
{ 'give': 1 },
{ 'it': 1 },
{ 'away': 1 },
{ 'to': 1 },
{ 'anyone': 1 },
{ 'you': 1 },
{ 'like': 1 },
{ 'but': 1 },
{ 'no': 1 },
{ 'charges': 1 },
{ 'are': 1 },
{ 'allowed': 1 },
{ 'welcome': 1 },
{ 'to': 1 },
{ 'the': 1 },
{ 'world': 1 },
{ 'of': 1 },
{ 'free': 1 },
{ 'plain': 1 },
{ 'vanilla': 1 },
{ 'electronic': 1 },
{ 'texts': 1 },

{'etexts': 1},
{'readable': 1},
{'by': 1},
{'both': 1},
{'humans': 1},
{'and': 1},
{'by': 1},
{'computers': 1},
{'since': 1},
{'1971': 1},
{'these': 1},
{'etexts': 1},
{'prepared': 1},
{'by': 1},
{'hundreds': 1},
{'of': 1},
{'volunteers': 1},
{'and': 1},
{'donations': 1},
{'information': 1},
{'on': 1},
{'contacting': 1},
{'project': 1},
{'gutenberg': 1},
{'to': 1},
{'get': 1},
{'etexts': 1},
{'and': 1},
{'further': 1},
{'information': 1},
{'is': 1},
{'included': 1},
{'below': 1},
{'we': 1},
{'need': 1},
{'your': 1},
{'donations': 1},
{'the': 1},
{'complete': 1},
{'works': 1},
{'of': 1},
{'william': 1},
{'shakespeare': 1},
{'january': 1},
{'1994': 1},
{'etext': 1},
{'100': 1},

{'the': 1},
{'library': 1},
{'of': 1},
{'the': 1},
{'future': 1},
{'complete': 1},
{'works': 1},
{'of': 1},
{'william': 1},
{'shakespeare': 1},
{'library': 1},
{'of': 1},
{'the': 1},
{'future': 1},
{'is': 1},
{'a': 1},
{'trademark': 1},
{'tm': 1},
{'of': 1},
{'world': 1},
{'library': 1},
{'inc': 1},
{'this': 1},
{'file': 1},
{'should': 1},
{'be': 1},
{'named': 1},
{'shaks12txt': 1},
{'or': 1},
{'shaks12zip': 1},
{'corrected': 1},
{'editions': 1},
{'of': 1},
{'our': 1},
{'etexts': 1},
{'get': 1},
{'a': 1},
{'new': 1},
{'number': 1},
{'shaks13txt': 1},
{'versions': 1},
{'based': 1},
{'on': 1},
{'separate': 1},
{'sources': 1},
{'get': 1},
{'new': 1},

{ 'letter': 1 },
{ 'shaks10atxt': 1 },
{ 'if': 1 },
{ 'you': 1 },
{ 'would': 1 },
{ 'like': 1 },
{ 'further': 1 },
{ 'information': 1 },
{ 'about': 1 },
{ 'world': 1 },
{ 'library': 1 },
{ 'inc': 1 },
{ 'please': 1 },
{ 'call': 1 },
{ 'them': 1 },
{ 'at': 1 },
{ '18004430238': 1 },
{ 'or': 1 },
{ 'email': 1 },
{ 'juliancnetcomcom': 1 },
{ 'please': 1 },
{ 'give': 1 },
{ 'them': 1 },
{ 'our': 1 },
{ 'thanks': 1 },
{ 'for': 1 },
{ 'their': 1 },
{ 'shakespeare': 1 },
{ 'cooperation': 1 },
{ 'the': 1 },
{ 'official': 1 },
{ 'release': 1 },
{ 'date': 1 },
{ 'of': 1 },
{ 'all': 1 },
{ 'project': 1 },
{ 'gutenberg': 1 },
{ 'etexts': 1 },
{ 'is': 1 },
{ 'at': 1 },
{ 'midnight': 1 },
{ 'central': 1 },
{ 'time': 1 },
{ 'of': 1 },
{ 'the': 1 },
{ 'last': 1 },
{ 'day': 1 },

{'of': 1},
{'the': 1},
{'stated': 1},
{'month': 1},
{'a': 1},
{'preliminary': 1},
{'version': 1},
{'may': 1},
{'often': 1},
{'be': 1},
{'posted': 1},
{'for': 1},
{'suggestion': 1},
{'comment': 1},
{'and': 1},
{'editing': 1},
{'by': 1},
{'those': 1},
{'who': 1},
{'wish': 1},
{'to': 1},
{'do': 1},
{'so': 1},
{'to': 1},
{'be': 1},
{'sure': 1},
{'you': 1},
{'have': 1},
{'an': 1},
{'up': 1},
{'to': 1},
{'date': 1},
{'first': 1},
{'edition': 1},
{'xxxxx10xxxx': 1},
{'please': 1},
{'check': 1},
{'file': 1},
{'sizes': 1},
{'in': 1},
{'the': 1},
{'first': 1},
{'week': 1},
{'of': 1},
{'the': 1},
{'next': 1},
{'month': 1},

{'since': 1},
{'our': 1},
{'ftp': 1},
{'program': 1},
{'has': 1},
{'a': 1},
{'bug': 1},
{'in': 1},
{'it': 1},
{'that': 1},
{'scrambles': 1},
{'the': 1},
{'date': 1},
{'tried': 1},
{'to': 1},
{'fix': 1},
{'and': 1},
{'failed': 1},
{'a': 1},
{'look': 1},
{'at': 1},
{'the': 1},
{'file': 1},
{'size': 1},
{'will': 1},
{'have': 1},
{'to': 1},
{'do': 1},
{'but': 1},
{'we': 1},
{'will': 1},
{'try': 1},
{'to': 1},
{'see': 1},
{'a': 1},
{'new': 1},
{'copy': 1},
{'has': 1},
{'at': 1},
{'least': 1},
{'one': 1},
{'byte': 1},
{'more': 1},
{'or': 1},
{'less': 1},
{'information': 1},
{'about': 1},

{'project': 1},
{'gutenberg': 1},
{'one': 1},
{'page': 1},
{'we': 1},
{'produce': 1},
{'about': 1},
{'two': 1},
{'million': 1},
{'dollars': 1},
{'for': 1},
{'each': 1},
{'hour': 1},
{'we': 1},
{'work': 1},
{'the': 1},
{'fifty': 1},
{'hours': 1},
{'is': 1},
{'one': 1},
{'conservative': 1},
{'estimate': 1},
{'for': 1},
{'how': 1},
{'long': 1},
{'it': 1},
{'we': 1},
{'take': 1},
{'to': 1},
{'get': 1},
{'any': 1},
{'etext': 1},
{'selected': 1},
{'entered': 1},
{'proofread': 1},
{'edited': 1},
{'copyright': 1},
{'searched': 1},
{'and': 1},
{'analyzed': 1},
{'the': 1},
{'copyright': 1},
{'letters': 1},
{'written': 1},
{'etc': 1},
{'this': 1},
{'projected': 1},

{'audience': 1},
{'is': 1},
{'one': 1},
{'hundred': 1},
{'million': 1},
{'readers': 1},
{'if': 1},
{'our': 1},
{'value': 1},
{'per': 1},
{'text': 1},
{'is': 1},
{'nominally': 1},
{'estimated': 1},
{'at': 1},
{'one': 1},
{'dollar': 1},
{'then': 1},
{'we': 1},
{'produce': 1},
{'2': 1},
{'million': 1},
{'dollars': 1},
{'per': 1},
{'hour': 1},
{'this': 1},
{'year': 1},
{'we': 1},
{'will': 1},
{'have': 1},
{'to': 1},
{'do': 1},
{'four': 1},
{'text': 1},
{'files': 1},
{'per': 1},
{'month': 1},
{'thus': 1},
{'upping': 1},
{'our': 1},
{'productivity': 1},
{'from': 1},
{'one': 1},
{'million': 1},
{'the': 1},
{'goal': 1},
{'of': 1},

{'project': 1},
{'gutenberg': 1},
{'is': 1},
{'to': 1},
{'give': 1},
{'away': 1},
{'one': 1},
{'trillion': 1},
{'etext': 1},
{'files': 1},
{'by': 1},
{'the': 1},
{'december': 1},
{'31': 1},
{'2001': 1},
{'10000': 1},
{'x': 1},
{'100000000trillion': 1},
{'this': 1},
{'is': 1},
{'ten': 1},
{'thousand': 1},
{'titles': 1},
{'each': 1},
{'to': 1},
{'one': 1},
{'hundred': 1},
{'million': 1},
{'readers': 1},
{'which': 1},
{'is': 1},
{'10': 1},
{'of': 1},
{'the': 1},
{'expected': 1},
{'number': 1},
{'of': 1},
{'computer': 1},
{'users': 1},
{'by': 1},
{'the': 1},
{'end': 1},
{'of': 1},
{'the': 1},
{'year': 1},
{'2001': 1},
{'we': 1},

{ 'need': 1 },
{ 'your': 1 },
{ 'donations': 1 },
{ 'more': 1 },
{ 'than': 1 },
{ 'ever': 1 },
{ 'all': 1 },
{ 'donations': 1 },
{ 'should': 1 },
{ 'be': 1 },
{ 'made': 1 },
{ 'to': 1 },
{ 'project': 1 },
{ 'gutenbergibc': 1 },
{ 'and': 1 },
{ 'are': 1 },
{ 'tax': 1 },
{ 'deductible': 1 },
{ 'to': 1 },
{ 'the': 1 },
{ 'extent': 1 },
{ 'allowable': 1 },
{ 'by': 1 },
{ 'law': 1 },
{ 'ibc': 1 },
{ 'is': 1 },
{ 'illinois': 1 },
{ 'benedictine': 1 },
{ 'college': 1 },
{ 'subscriptions': 1 },
{ 'to': 1 },
{ 'our': 1 },
{ 'paper': 1 },
{ 'newsletter': 1 },
{ 'go': 1 },
{ 'to': 1 },
{ 'ibc': 1 },
{ 'too': 1 },
{ 'for': 1 },
{ 'these': 1 },
{ 'and': 1 },
{ 'other': 1 },
{ 'matters': 1 },
{ 'please': 1 },
{ 'mail': 1 },
{ 'to': 1 },
{ 'project': 1 },

{'gutenberg': 1},
{'p': 1},
{'o': 1},
{'box': 1},
{'2782': 1},
{'champaign': 1},
{'il': 1},
{'61825': 1},
{'when': 1},
{'all': 1},
{'other': 1},
{'email': 1},
{'fails': 1},
{'try': 1},
{'our': 1},
{'michael': 1},
{'s': 1},
{'hart': 1},
{'executive': 1},
{'director': 1},
{'hartvmdcsouiucedu': 1},
{'internet': 1},
{'hartuiucvmd': 1},
{'bitnet': 1},
{'we': 1},
{'would': 1},
{'prefer': 1},
{'to': 1},
{'send': 1},
{'you': 1},
{'this': 1},
{'information': 1},
{'by': 1},
{'email': 1},
{'internet': 1},
{'bitnet': 1},
{'compuserve': 1},
{'attmail': 1},
{'or': 1},
{'mcimail': 1},
{'': 1},
{'if': 1},
{'you': 1},
{'have': 1},
{'an': 1},
{'ftp': 1},
{'program': 1},

{'or': 1},
{'emulator': 1},
{'please': 1},
{'ftp': 1},
{'directly': 1},
{'to': 1},
{'the': 1},
{'project': 1},
{'gutenberg': 1},
{'archives': 1},
{'mac': 1},
{'users': 1},
{'do': 1},
{'not': 1},
{'point': 1},
{'and': 1},
{'click': 1},
{'': 1},
{'type': 1},
{'ftp': 1},
{'mrcnextcsouiucedu': 1},
{'login': 1},
{'anonymous': 1},
{'password': 1},
{'yourlogin': 1},
{'cd': 1},
{'etextetext91': 1},
{'or': 1},
{'cd': 1},
{'etext92': 1},
{'or': 1},
{'cd': 1},
{'etext93': 1},
{'for': 1},
{'new': 1},
{'books': 1},
{'now': 1},
{'also': 1},
{'in': 1},
{'cd': 1},
{'etextetext93': 1},
{'or': 1},
{'cd': 1},
{'etextarticles': 1},
{'get': 1},
{'suggest': 1},
{'gut': 1},

{'for': 1},
{'more': 1},
{'information': 1},
{'dir': 1},
{'to': 1},
{'see': 1},
{'files': 1},
{'get': 1},
{'or': 1},
{'mget': 1},
{'to': 1},
{'get': 1},
{'files': 1},
{'': 1},
{'set': 1},
{'bin': 1},
{'for': 1},
{'zip': 1},
{'files': 1},
{'get': 1},
{'0indexgut': 1},
{'for': 1},
{'a': 1},
{'list': 1},
{'of': 1},
{'books': 1},
{'and': 1},
{'get': 1},
{'new': 1},
{'gut': 1},
{'for': 1},
{'general': 1},
{'information': 1},
{'and': 1},
{'mget': 1},
{'gut': 1},
{'for': 1},
{'newsletters': 1},
{'information': 1},
{'prepared': 1},
{'by': 1},
{'the': 1},
{'project': 1},
{'gutenberg': 1},
{'legal': 1},
{'advisor': 1},
{'': 1},

{'small': 1},
{'print': 1},
{'for': 1},
{'complete': 1},
{'shakespeare': 1},
{'': 1},
{'this': 1},
{'electronic': 1},
{'version': 1},
{'of': 1},
{'the': 1},
{'complete': 1},
{'works': 1},
{'of': 1},
{'william': 1},
{'shakespeare': 1},
{'is': 1},
{'copyright': 1},
{'19901993': 1},
{'by': 1},
{'world': 1},
{'library': 1},
{'inc': 1},
{'and': 1},
{'is': 1},
{'provided': 1},
{'by': 1},
{'project': 1},
{'gutenberg': 1},
{'etext': 1},
{'of': 1},
{'illinois': 1},
{'benedictine': 1},
{'college': 1},
{'with': 1},
{'permission': 1},
{'since': 1},
{'unlike': 1},
{'many': 1},
{'other': 1},
{'project': 1},
{'gutenbergtm': 1},
{'etexts': 1},
{'this': 1},
{'etext': 1},
{'is': 1},
{'copyright': 1},

{'protected': 1},
{'and': 1},
{'since': 1},
{'the': 1},
{'materials': 1},
{'and': 1},
{'methods': 1},
{'you': 1},
{'use': 1},
{'will': 1},
{'effect': 1},
{'the': 1},
{'projects': 1},
{'reputation': 1},
{'your': 1},
{'right': 1},
{'to': 1},
{'copy': 1},
{'and': 1},
{'distribute': 1},
{'it': 1},
{'is': 1},
{'limited': 1},
{'by': 1},
{'the': 1},
{'copyright': 1},
{'and': 1},
{'other': 1},
{'laws': 1},
{'and': 1},
{'by': 1},
{'the': 1},
{'conditions': 1},
{'of': 1},
{'this': 1},
{'small': 1},
{'print': 1},
{'statement': 1},
{'1': 1},
{'license': 1},
{'a': 1},
{'you': 1},
{'may': 1},
{'and': 1},
{'are': 1},
{'encouraged': 1},
{'to': 1},

{'distribute': 1},
{'electronic': 1},
{'and': 1},
{'machine': 1},
{'readable': 1},
{'copies': 1},
{'of': 1},
{'this': 1},
{'etext': 1},
{'so': 1},
{'long': 1},
{'as': 1},
{'such': 1},
{'copies': 1},
{'1': 1},
{'are': 1},
{'for': 1},
{'your': 1},
{'or': 1},
{'others': 1},
{'personal': 1},
{'use': 1},
{'only': 1},
{'and': 1},
{'2': 1},
{'are': 1},
{'not': 1},
{'distributed': 1},
{'or': 1},
{'used': 1},
{'commercially': 1},
{'prohibited': 1},
{'commercial': 1},
{'distribution': 1},
{'includes': 1},
{'by': 1},
{'any': 1},
{'service': 1},
{'that': 1},
{'charges': 1},
{'for': 1},
{'download': 1},
{'time': 1},
{'or': 1},
{'for': 1},
{'membership': 1},
{'b': 1},

{ 'this': 1 },
{ 'license': 1 },
{ 'is': 1 },
{ 'subject': 1 },
{ 'to': 1 },
{ 'the': 1 },
{ 'conditions': 1 },
{ 'that': 1 },
{ 'you': 1 },
{ 'honor': 1 },
{ 'the': 1 },
{ 'refund': 1 },
{ 'and': 1 },
{ 'replacement': 1 },
{ 'provisions': 1 },
{ 'of': 1 },
{ 'this': 1 },
{ 'small': 1 },
{ 'print': 1 },
{ 'statement': 1 },
{ 'and': 1 },
{ 'that': 1 },
{ 'you': 1 },
{ 'distribute': 1 },
{ 'exact': 1 },
{ 'copies': 1 },
{ 'of': 1 },
{ 'this': 1 },
{ 'etext': 1 },
{ 'including': 1 },
{ 'this': 1 },
{ 'small': 1 },
{ 'print': 1 },
{ 'statement': 1 },
{ 'such': 1 },
{ 'copies': 1 },
{ 'can': 1 },
{ 'be': 1 },
{ 'compressed': 1 },
{ 'or': 1 },
{ 'any': 1 },
{ 'proprietary': 1 },
{ 'form': 1 },
{ 'including': 1 },
{ 'any': 1 },
{ 'form': 1 },
{ 'resulting': 1 },


```

{'from': 1},
{'word': 1},
{'processing': 1},
{'or': 1},
{'hypertext': 1},
{'software': 1},
{'so': 1},
{'long': 1},
{'as': 1},
{'either': 1},
{'1': 1},
{'the': 1},
{'etext': 1},
{'when': 1},
{'displayed': 1},
{'is': 1},
{'clearly': 1},
{'readable': 1},
{'and': 1},
{'does': 1},
{'not': 1},
{'contain': 1},
{'characters': 1},
{'other': 1},
{'than': 1},
{'those': 1},
{'intended': 1},
{'by': 1},
{'the': 1},
{'author': 1},
{'of': 1},
{'the': 1},
{'work': 1},
{'although': 1},
{'tilde': 1},
{'': 1},
{'asterisk': 1},
...]
```

```
[10]: def reducer(i, j):
        for k in j: i[k] = i.get(k, 0) + j.get(k, 0)
        return i
```

```
[11]: reduce(reducer, map_lst)
```

```
[11]: {'this': 6847,
       'is': 9600,
```

'the': 27643,
'100th': 1,
'etext': 242,
'file': 19,
'presented': 18,
'by': 4412,
'project': 249,
'gutenberg': 233,
'and': 26730,
'in': 10967,
'cooperation': 2,
'with': 7996,
'world': 849,
'library': 232,
'inc': 224,
'from': 2646,
'their': 2077,
'of': 18173,
'future': 17,
'shakespeare': 268,
'cdroms': 1,
'often': 123,
'releases': 1,
'etexts': 8,
'that': 11121,
'are': 3851,
'not': 8725,
'placed': 10,
'public': 57,
'domain': 4,
'has': 389,
'certain': 176,
'copyright': 228,
'implications': 1,
'you': 13649,
'should': 1576,
'read': 202,
'electronic': 443,
'version': 222,
'complete': 246,
'works': 251,
'william': 348,
'19901993': 221,
'provided': 252,
'illinois': 223,
'benedictine': 223,
'college': 226,

'permission': 225,
'machine': 223,
'readable': 224,
'copies': 446,
'may': 1865,
'be': 7090,
'distributed': 443,
'so': 5268,
'long': 671,
'as': 5960,
'such': 1615,
'1': 308,
'for': 8244,
'your': 6882,
'or': 3123,
'others': 449,
'personal': 234,
'use': 548,
'only': 532,
'2': 301,
'used': 242,
'commercially': 221,
'prohibited': 221,
'commercial': 221,
'distribution': 224,
'includes': 223,
'any': 936,
'service': 443,
'charges': 237,
'download': 221,
'time': 1270,
'membership': 221,
'proud': 203,
'to': 19198,
'cooperate': 1,
'presentation': 4,
'reading': 34,
'education': 9,
'entertainment': 37,
'however': 13,
'neither': 185,
'shareware': 1,
'nor': 969,
'': 338,
'under': 290,
'conditions': 29,
'no': 3796,

'made': 827,
'access': 23,
'material': 6,
'encouraged': 3,
'give': 1333,
'it': 7690,
'away': 858,
'anyone': 2,
'like': 1703,
'but': 6270,
'allowed': 5,
'welcome': 365,
'free': 187,
'plain': 115,
'vanilla': 1,
'texts': 1,
'both': 641,
'humans': 1,
'computers': 1,
'since': 448,
'1971': 1,
'these': 1324,
'prepared': 12,
'hundreds': 4,
'volunteers': 1,
'donations': 4,
'information': 10,
'on': 3051,
'contacting': 1,
'get': 312,
'further': 190,
'included': 2,
'below': 54,
'we': 3306,
'need': 180,
'january': 3,
'1994': 1,
'100': 2,
'a': 14613,
'trademark': 1,
'tm': 1,
'named': 5,
'shaks12txt': 1,
'shaks12zip': 1,
'corrected': 3,
'editions': 1,
'our': 3068,

'new': 213,
'number': 51,
'shaks13txt': 1,
'versions': 1,
'based': 1,
'separate': 5,
'sources': 2,
'letter': 245,
'shaks10atxt': 1,
'if': 3503,
'would': 2295,
'about': 403,
'please': 393,
'call': 555,
'them': 1980,
'at': 2509,
'18004430238': 1,
'email': 3,
'juliancnetcomcom': 1,
'thanks': 172,
'official': 3,
'release': 8,
'date': 22,
'all': 3887,
'midnight': 43,
'central': 1,
'last': 268,
'day': 720,
'stated': 1,
'month': 42,
'preliminary': 1,
'posted': 7,
'suggestion': 10,
'comment': 14,
'editing': 1,
'those': 547,
'who': 1203,
'wish': 209,
'do': 3754,
'sure': 292,
'have': 5887,
'an': 1836,
'up': 1069,
'first': 1161,
'edition': 2,
'xxxxx10xxxx': 1,
'check': 28,

'sizes': 3,
'week': 29,
'next': 167,
'ftp': 4,
'program': 3,
'bug': 3,
'scrambles': 1,
'tried': 20,
'fix': 4,
'failed': 1,
'look': 829,
'size': 13,
'will': 4979,
'try': 93,
'see': 1440,
'copy': 16,
'least': 101,
'one': 1787,
'byte': 1,
'more': 2291,
'less': 221,
'page': 429,
'produce': 19,
'two': 703,
'million': 19,
'dollars': 3,
'each': 240,
'hour': 299,
'work': 186,
'fifty': 31,
'hours': 150,
'conservative': 1,
'estimate': 7,
'how': 2160,
'take': 1198,
'selected': 1,
'entered': 3,
'proofread': 1,
'edited': 1,
'searched': 1,
'analyzed': 1,
'letters': 129,
'written': 35,
'etc': 19,
'projected': 1,
'audience': 33,
'hundred': 114,

'readers': 2,
'value': 27,
'per': 6,
'text': 13,
'nominally': 1,
'estimated': 1,
'dollar': 2,
'then': 2222,
'year': 92,
'four': 149,
'files': 9,
'thus': 782,
'upping': 1,
'productivity': 1,
'goal': 4,
'trillion': 1,
'december': 7,
'31': 2,
'2001': 2,
'10000': 1,
'x': 5,
'1000000000trillion': 1,
'ten': 155,
'thousand': 322,
'titles': 24,
'which': 2316,
'10': 4,
'expected': 11,
'computer': 2,
'users': 2,
'end': 345,
'than': 1881,
'ever': 596,
'gutenbergibc': 1,
'tax': 13,
'deductible': 1,
'extent': 6,
'allowable': 1,
'law': 192,
'ibc': 2,
'subscriptions': 1,
'paper': 104,
'newsletter': 1,
'go': 1734,
'too': 1233,
'other': 678,
'matters': 28,

'mail': 4,
'p': 9,
'o': 2608,
'box': 26,
'2782': 2,
'champaign': 2,
'il': 13,
'61825': 2,
'when': 2051,
'fails': 11,
'michael': 27,
's': 44,
'hart': 11,
'executive': 1,
'director': 1,
'hartvmcdsouiucedu': 2,
'internet': 4,
'hartuiucvmd': 2,
'bitnet': 3,
'prefer': 20,
'send': 258,
'compuserve': 2,
'attmail': 2,
'mcimail': 1,
'emulator': 2,
'directly': 34,
'archives': 1,
'mac': 1,
'point': 143,
'click': 1,
'type': 3,
'mrcnextcsouiucedu': 1,
'login': 1,
'anonymous': 1,
'password': 1,
'yourlogin': 1,
'cd': 5,
'etextetext91': 1,
'etext92': 1,
'etext93': 1,
'books': 38,
'now': 2779,
'also': 41,
'etextetext93': 1,
'etextarticles': 1,
'suggest': 8,
'gut': 3,

'dir': 1,
'mget': 2,
'set': 462,
'bin': 4,
'zip': 1,
'Oindexgut': 1,
'list': 58,
'general': 204,
'newsletters': 1,
'legal': 4,
'advisor': 1,
'small': 99,
'print': 22,
'unlike': 16,
'many': 572,
'gutenbergtm': 2,
'protected': 4,
'materials': 1,
'methods': 2,
'effect': 72,
'projects': 3,
'reputation': 47,
'right': 350,
'distribute': 5,
'limited': 8,
'laws': 41,
'statement': 3,
'license': 9,
'b': 12,
'subject': 102,
'honor': 36,
'refund': 3,
'replacement': 3,
'provisions': 1,
'exact': 9,
'including': 6,
'can': 1213,
'compressed': 1,
'proprietary': 1,
'form': 113,
'resulting': 1,
'word': 497,
'processing': 1,
'hypertext': 2,
'software': 2,
'either': 166,
'displayed': 2,

'clearly': 6,
'does': 324,
'contain': 17,
'characters': 13,
'intended': 20,
'author': 15,
'although': 79,
'tilde': 1,
'asterisk': 1,
'underline': 1,
'convey': 34,
'punctuation': 1,
'additional': 2,
'indicate': 1,
'links': 3,
'readily': 2,
'convertible': 1,
'reader': 2,
'expense': 19,
'into': 640,
'ascii': 2,
'ebcdic': 1,
'equivalent': 1,
'displays': 1,
'case': 109,
'instance': 26,
'most': 1182,
'processors': 1,
'3': 46,
'provide': 32,
'agree': 21,
'request': 65,
'cost': 48,
'fee': 27,
'warranty': 5,
'disclaimer': 1,
'damages': 5,
'defect': 15,
'incomplete': 1,
'inaccurate': 1,
'corrupt': 30,
'data': 4,
'transcription': 1,
'errors': 14,
'infringement': 1,
'defective': 6,
'damaged': 1,

'disk': 1,
'virus': 1,
'codes': 1,
'damage': 4,
'cannot': 762,
'equipment': 1,
'described': 2,
'party': 50,
'receive': 96,
'disclaims': 2,
'liability': 2,
'costs': 4,
'expenses': 3,
'fees': 6,
'remedies': 9,
'negligence': 15,
'strict': 17,
'breach': 32,
'contract': 27,
'indirect': 9,
'consequential': 1,
'punitive': 1,
'incidental': 1,
'even': 567,
'notice': 31,
'possibility': 4,
'discover': 35,
'within': 493,
'90': 2,
'days': 230,
'receiv': 1,
'ing': 1,
'money': 182,
'paid': 65,
'sending': 12,
'explanatory': 1,
'note': 150,
'person': 149,
'received': 28,
'physical': 3,
'medium': 2,
'must': 1492,
'return': 196,
'choose': 105,
'alternatively': 2,
'electronically': 2,
'second': 572,

'opportunity': 16,
'otherwise': 37,
'asis': 1,
'warranties': 3,
'kind': 278,
'express': 36,
'implied': 2,
'merchantability': 1,
'fitness': 10,
'particular': 56,
'purpose': 207,
'some': 1338,
'states': 19,
'allow': 30,
'disclaimers': 2,
'exclusion': 1,
'limitation': 3,
'consequen': 1,
'tial': 1,
'above': 147,
'exclusions': 1,
'apply': 16,
'rights': 22,
'indemnity': 1,
'indemnify': 1,
'hold': 461,
'its': 44,
'directors': 1,
'officers': 87,
'members': 11,
'agents': 4,
'harmless': 16,
'lia': 1,
'bility': 1,
'arise': 38,
'indirectly': 9,
'following': 52,
'cause': 337,
'alteration': 10,
'modification': 1,
'addition': 28,
'c': 21,
'4': 32,
'what': 4465,
'want': 134,
'dont': 1,
'dedicated': 4,

'increasing': 8,
'licensed': 1,
'freely': 55,
'gratefully': 1,
'accepts': 3,
'contributions': 1,
'scanning': 1,
'machines': 1,
'ocr': 1,
'royalty': 25,
'licenses': 1,
'whatever': 10,
'else': 416,
'think': 1025,
'pro': 2,
'ject': 1,
'association': 1,
'write': 125,
'us': 1622,
'reached': 1,
'internethartvmdcsouiucedu': 1,
'internetvmdcsouiuceduhart': 1,
'prof': 1,
'po': 2,
'charles': 126,
'kramer': 1,
'attorney': 8,
'726002026compuservecom': 1,
'tel': 1,
'2122545093': 1,
'v120893': 1,
'1609': 3,
'sonnets': 5,
'fairest': 41,
'creatures': 41,
'desire': 218,
'increase': 28,
'thereby': 26,
'beautys': 30,
'rose': 64,
'might': 496,
'never': 1011,
'die': 474,
'riper': 3,
'decease': 6,
'his': 6857,
'tender': 138,

'heir': 108,
'bear': 534,
'memory': 58,
'thou': 5485,
'contracted': 11,
'thine': 455,
'own': 771,
'bright': 68,
'eyes': 690,
'feedst': 3,
'thy': 4032,
'lights': 28,
'flame': 23,
'selfsubstantial': 1,
'fuel': 5,
'making': 81,
'famine': 12,
'where': 1232,
'abundance': 13,
'lies': 245,
'self': 138,
'foe': 62,
'sweet': 789,
'cruel': 70,
'art': 915,
'worlds': 53,
'fresh': 87,
'ornament': 18,
'herald': 58,
'gaudy': 6,
'spring': 59,
'bud': 14,
'buriest': 1,
'content': 157,
'churl': 9,
'makst': 22,
'waste': 44,
'niggarding': 1,
'pity': 217,
'glutton': 3,
'eat': 142,
'due': 76,
'grave': 177,
'thee': 3178,
'forty': 28,
'winters': 33,
'shall': 3591,

'besiege': 10,
'brow': 69,
'dig': 10,
'deep': 136,
'trenches': 8,
'field': 187,
'youths': 10,
'livery': 21,
'gazed': 1,
'tattered': 3,
'weed': 18,
'worth': 204,
'held': 84,
'being': 662,
'asked': 4,
'beauty': 201,
'treasure': 51,
'lusty': 34,
'say': 1680,
'sunken': 3,
'were': 1577,
'alleating': 1,
'shame': 302,
'thriftless': 4,
'praise': 182,
'much': 1035,
'deserved': 19,
'couldst': 30,
'answer': 380,
'fair': 757,
'child': 232,
'mine': 1161,
'sum': 56,
'my': 12481,
'count': 107,
'make': 1629,
'old': 711,
'excuse': 78,
'proving': 1,
'succession': 17,
'blood': 639,
'warm': 44,
'feelst': 1,
'cold': 195,
'glass': 56,
'tell': 1059,
'face': 416,

'viewest': 2,
'another': 430,
'whose': 615,
'repair': 45,
'renewest': 1,
'dost': 429,
'beguile': 24,
'unbless': 1,
'mother': 360,
'she': 2208,
'uneared': 1,
'womb': 53,
'disdains': 6,
'tillage': 1,
'husbandry': 14,
'he': 6251,
'fond': 56,
'tomb': 54,
'selflove': 7,
'stop': 95,
'posterity': 11,
'mothers': 93,
'calls': 92,
'back': 335,
'lovely': 53,
'april': 15,
'her': 3843,
'prime': 17,
'through': 261,
'windows': 19,
'age': 197,
'shalt': 295,
'despite': 52,
'wrinkles': 14,
'golden': 91,
'live': 513,
'remembered': 5,
'single': 58,
'image': 46,
'dies': 102,
'unthrifty': 5,
'loveliness': 2,
'why': 1465,
'spend': 66,
'upon': 1731,
'legacy': 5,
'natures': 66,

'bequest': 1,
'gives': 147,
'nothing': 633,
'doth': 939,
'lend': 93,
'frank': 11,
'lends': 13,
'beauteous': 39,
'niggard': 8,
'abuse': 46,
'bounteous': 18,
'largess': 5,
'given': 191,
'profitless': 3,
'usurer': 9,
'great': 899,
'sums': 17,
'yet': 1569,
'canst': 173,
'having': 155,
'traffic': 10,
'alone': 236,
'deceive': 28,
'nature': 344,
'gone': 464,
'acceptable': 1,
'audit': 9,
'leave': 651,
'unused': 5,
'tomed': 1,
'lives': 204,
'th': 1148,
'executor': 2,
'5': 20,
'gentle': 366,
'did': 1626,
'frame': 46,
'gaze': 27,
'every': 491,
'eye': 418,
'dwell': 39,
'play': 303,
'tyrants': 20,
'very': 813,
'same': 218,
'unfair': 1,
'fairly': 46,

'excel': 6,
'neverresting': 1,
'leads': 32,
'summer': 54,
'hideous': 22,
'winter': 47,
'confounds': 11,
'him': 5192,
'there': 1808,
'sap': 12,
'checked': 2,
'frost': 12,
'leaves': 70,
'quite': 86,
'oersnowed': 1,
'bareness': 4,
'summers': 37,
'distillation': 2,
'left': 245,
'liquid': 9,
'prisoner': 81,
'pent': 8,
'walls': 85,
'bereft': 17,
'remembrance': 63,
'was': 2229,
'flowers': 47,
'distilled': 5,
'though': 628,
'they': 2471,
'meet': 324,
'leese': 1,
'show': 416,
'substance': 41,
'still': 552,
'6': 10,
'let': 2099,
'ragged': 23,
'hand': 838,
'deface': 4,
'ere': 359,
'vial': 5,
'place': 430,
'selfkilled': 1,
'forbidden': 5,
'usury': 4,
'happies': 1,

'pay': 138,
'willing': 39,
'loan': 3,
'thats': 399,
'breed': 49,
'times': 236,
'happier': 17,
'refigured': 1,
'could': 623,
'death': 858,
'shouldst': 75,
'depart': 65,
'leaving': 25,
'living': 120,
'selfwilled': 1,
'deaths': 50,
'conquest': 31,
'worms': 23,
'7': 7,
'lo': 74,
'orient': 5,
'gracious': 195,
'light': 264,
'lifts': 5,
'burning': 45,
'head': 514,
'homage': 17,
'newappearing': 2,
'sight': 199,
'serving': 7,
'looks': 214,
'sacred': 47,
'majesty': 256,
'climbed': 1,
'steepup': 1,
'heavenly': 54,
'hill': 40,
'resembling': 4,
'strong': 185,
'youth': 277,
'middle': 21,
'mortal': 96,
'adore': 14,
'attending': 43,
'pilgrimage': 16,
'highmost': 2,
'pitch': 32,

'weary': 76,
'car': 35,
'feeble': 37,
'reeleth': 1,
'fore': 31,
'duteous': 11,
'converted': 7,
'low': 96,
'tract': 4,
'way': 562,
'outgoing': 1,
'noon': 17,
'unlooked': 2,
'diest': 22,
'unless': 153,
'son': 691,
'8': 3,
'music': 201,
'hear': 871,
'hearst': 10,
'sadly': 18,
'sweets': 16,
'war': 252,
'joy': 191,
'delights': 20,
'lovst': 26,
'receivst': 2,
'gladly': 16,
'pleasure': 179,
'annoy': 8,
'true': 792,
'concord': 9,
'welltuned': 1,
'sounds': 47,
'unions': 1,
'married': 149,
'offend': 59,
'ear': 191,
'sweetly': 27,
'chide': 49,
'singleness': 2,
'parts': 112,
'mark': 226,
'string': 6,
'husband': 291,
'strikes': 47,
'mutual': 15,

'ordering': 3,
'sire': 18,
'happy': 202,
'pleasing': 22,
'sing': 139,
'speechless': 13,
'song': 79,
'seeming': 57,
'sings': 91,
'wilt': 307,
'prove': 252,
'none': 467,
'9': 3,
'fear': 625,
'wet': 19,
'widows': 19,
'consumst': 1,
'life': 834,
'ah': 183,
'issueless': 2,
'hap': 26,
'wail': 23,
'makeless': 1,
'wife': 510,
'widow': 110,
'weep': 163,
'hast': 584,
'behind': 118,
'private': 66,
'well': 2462,
'keep': 477,
'childrens': 13,
'husbands': 63,
'shape': 87,
'mind': 357,
'unthrift': 3,
'shifts': 7,
'enjoys': 7,
'hath': 1941,
'kept': 97,
'user': 1,
'destroys': 2,
'love': 2053,
'toward': 103,
'bosom': 125,
'sits': 59,
'himself': 451,

'murdrous': 11,
'commits': 5,
'deny': 133,
'bearst': 7,
'unprovident': 1,
'grant': 109,
'beloved': 29,
'evident': 7,
'possessed': 5,
'hate': 177,
'gainst': 101,
'stickst': 3,
'conspire': 8,
'seeking': 24,
'roof': 16,
'ruinate': 2,
'chief': 108,
'change': 155,
'thought': 375,
'i': 20681,
'fairer': 38,
'lodged': 5,
'presence': 106,
'kindhearted': 1,
'me': 7768,
'11': 1,
'fast': 129,
'wane': 2,
'growst': 3,
'departest': 1,
'youngly': 2,
'bestowst': 1,
'mayst': 68,
'convertest': 1,
'herein': 29,
'wisdom': 93,
'without': 364,
'folly': 79,
'decay': 24,
'minded': 10,
'cease': 47,
'threescore': 7,
'whom': 423,
'store': 38,
'harsh': 27,
'featureless': 1,
'rude': 71,

```
'barrenly': 1,  
'perish': 31,  
'best': 459,  
'endowed': 2,  
'gave': 229,  
'gift': 65,  
'bounty': 40,  
'cherish': 18,  
'carved': 6,  
'seal': 66,  
'meant': 49,  
...}
```

[]:

[]: