DATA 603

Name : Akash Reddy Loka

Campus ID : LB04019

Assignment 2 : MapReduce Assignment

```python
In [34]:  import re
          import contractions
          import numpy as np
```

Birth day - 15 Nov 2000

Month - 11

Date - 15

Year - 2000

```python
In [51]:  with open('file1.txt','r',encoding='utf-8') as file:
              text1 = file.read()
```

```python
In [52]:  with open('file2.txt','r',encoding='utf-8') as file:
              text2 = file.read()
```

```python
In [53]:  #http://www.gwicks.net/dictionaries.htm
          english_words = []
          with open('english3.txt','r') as file:
              for word in file:
                  english_words.append(word.strip())
```

```python
In [54]:  def parsing(text):
              """
              This function performs text pre-processing
              and returns clean text
              """

              text = contractions.fix(text) # expands all the contractions in the text
              text = re.sub(r'[^\w\s]', '', text) # removes all the punctuation
              text = text.lower() # converts the text to lower case
              text = text.replace('\n',' ')  #replace newline character with empty string
              text = text.strip() # removes leading and trailing edge spaces of the strir
              text = re.sub(' +',' ',text) # replaces more than one spaces with single sp

              return text
```

```python
In [55]:  parsed_text1 = parsing(text1)
          parsed_text2 = parsing(text2)
```

```python
In [56]:    def map_word_counts(text):
                """
                This functions returns tokenized words list from the given text
                """
                words = text.split(' ')
                return words
```

```python
In [57]:    def reducer(words_list):
                """
                This functions returns word counts for each unique word in the words list
                in the form of dictionary with key being word and value being counts.
                """
                unique, counts = np.unique(words_list, return_counts=True)
                word_counts = {}
                for i,j in zip(unique,counts):
                    word_counts[i] = j
                return word_counts
```

```python
In [58]:    words = map_word_counts(parsed_text1)
            words
```

```
Out[58]:  ['amelia',
           'bones',
           'head',
           'of',
           'the',
           'department',
           'of',
           'magical',
           'law',
           'enforcement',
           'we',
           'think',
           'hewhomustnotbenamed',
           'may',
           'have',
           'murdered',
           'her',
           'in',
           'person',
           'because',
           'she',
           'was',
           'a',
           'very',
           'gifted',
           'witch',
           'and',
           'and',
           'all',
           'the',
           'evidence',
           'was',
           'that',
           'she',
           'put',
           'up',
           'a',
           'real',
           'fight',
           'fudge',
           'cleared',
           'his',
           'throat',
           'and',
           'with',
           'an',
           'effort',
           'it',
           'seemed',
           'stopped',
           'spinning',
           'his',
           'bowler',
           'hat',
           'but',
           'that',
           'murder',
           'was',
           'in',
           'the',
```

'newspapers',
'said',
'the',
'prime',
'minister',
'momentarily',
'diverted',
'from',
'his',
'anger',
'our',
'newspapers',
'amelia',
'bones',
'it',
'just',
'said',
'she',
'was',
'a',
'middleaged',
'woman',
'who',
'lived',
'alone',
'it',
'was',
'a',
'a',
'nasty',
'killing',
'was',
'not',
'it',
'it',
'is',
'had',
'rather',
'a',
'lot',
'of',
'publicity',
'the',
'police',
'are',
'baffled',
'you',
'see',
'fudge',
'sighed',
'well',
'of',
'course',
'they',
'are',
'he',
'said',
'killed',
'in',
'a',

'room',
'that',
'was',
'locked',
'from',
'the',
'inside',
'was',
'not',
'she',
'we',
'on',
'the',
'other',
'hand',
'know',
'exactly',
'who',
'did',
'it',
'not',
'that',
'that',
'gets',
'us',
'any',
'further',
'toward',
'catching',
'him',
'and',
'then',
'there',
'was',
'emmeline',
'vance',
'maybe',
'you',
'did',
'not',
'hear',
'about',
'that',
'one',
'oh',
'yes',
'i',
'did',
'said',
'the',
'prime',
'minister',
'it',
'happened',
'just',
'around',
'the',
'corner',
'from',
'here',

'as',
'a',
'matter',
'of',
'fact',
'the',
'papers',
'had',
'a',
'field',
'day',
'with',
'it',
'breakdown',
'of',
'law',
'and',
'order',
'in',
'the',
'prime',
'ministers',
'backyard',
'and',
'as',
'if',
'all',
'that',
'was',
'not',
'enough',
'said',
'fudge',
'barely',
'listening',
'to',
'the',
'prime',
'minister',
'we',
'have',
'got',
'dementors',
'swarming',
'all',
'over',
'the',
'place',
'attacking',
'people',
'left',
'right',
'and',
'center',
'once',
'upon',
'a',
'happier',
'time',
'this',

'sentence',
'would',
'have',
'been',
'unintelligible',
'to',
'the',
'prime',
'minister',
'but',
'he',
'was',
'wiser',
'now',
'i',
'thought',
'dementors',
'guard',
'the',
'prisoners',
'in',
'azkaban',
'he',
'said',
'cautiously',
'they',
'did',
'said',
'fudge',
'wearily',
'but',
'not',
'anymore',
'they',
'have',
'deserted',
'the',
'prison',
'and',
'joined',
'hewhomustnotbenamed',
'i',
'will',
'not',
'pretend',
'that',
'was',
'not',
'a',
'blow',
'but',
'said',
'the',
'prime',
'minister',
'with',
'a',
'sense',
'of',
'dawning',

'horror',
'did',
'not',
'you',
'tell',
'me',
'they',
'are',
'the',
'creatures',
'that',
'drain',
'hope',
'and',
'happiness',
'out',
'of',
'people',
'that',
'is',
'right',
'and',
'they',
'are',
'breeding',
'that',
'is',
'what',
'is',
'causing',
'all',
'this',
'mist',
'the',
'prime',
'minister',
'sank',
'weakkneed',
'into',
'the',
'nearest',
'chair',
'the',
'idea',
'of',
'invisible',
'creatures',
'swooping',
'through',
'the',
'towns',
'and',
'countryside',
'spreading',
'despair',
'and',
'hopelessness',
'in',
'his',
'voters',

'made',
'him',
'feel',
'quite',
'faint',
'now',
'see',
'here',
'fudge',
'you',
'have',
'got',
'to',
'do',
'something',
'it',
'is',
'your',
'responsibility',
'as',
'minister',
'of',
'magic',
'my',
'dear',
'prime',
'minister',
'you',
'cannot',
'honestly',
'think',
'i',
'am',
'still',
'minister',
'of',
'magic',
'after',
'all',
'this',
'i',
'was',
'sacked',
'three',
'days',
'ago',
'the',
'whole',
'wizarding',
'community',
'has',
'been',
'screaming',
'for',
'my',
'resignation',
'for',
'a',
'fortnight',
'i',

'have',
'never',
'known',
'them',
'so',
'united',
'in',
'my',
'whole',
'term',
'of',
'office',
'said',
'fudge',
'with',
'a',
'brave',
'attempt',
'at',
'a',
'smile',
'the',
'prime',
'minister',
'was',
'momentarily',
'lost',
'for',
'words',
'despite',
'his',
'indignation',
'at',
'the',
'position',
'into',
'which',
'he',
'had',
'been',
'placed',
'he',
'still',
'rather',
'felt',
'for',
'the',
'shrunkenlooking',
'man',
'sitting',
'opposite',
'him',
'i',
'am',
'very',
'sorry',
'he',
'said',
'finally',
'if',

'there',
'is',
'anything',
'i',
'can',
'do',
'it',
'is',
'very',
'kind',
'of',
'you',
'prime',
'minister',
'but',
'there',
'is',
'nothing',
'i',
'was',
'sent',
'here',
'tonight',
'to',
'bring',
'you',
'up',
'to',
'date',
'on',
'recent',
'events',
'and',
'to',
'introduce',
'you',
'to',
'my',
'successor',
'i',
'rather',
'thought',
'he',
'would',
'be',
'here',
'by',
'now',
'but',
'of',
'course',
'he',
'is',
'very',
'busy',
'at',
'the',
'moment',
'with',
'so',

'much',
'going',
'on',
'fudge',
'looked',
'around',
'at',
'the',
'portrait',
'of',
'the',
'ugly',
'little',
'man',
'wearing',
'the',
'long',
'curly',
'silver',
'wig',
'who',
'was',
'digging',
'in',
'his',
'ear',
'with',
'the',
'point',
'of',
'a',
'quill',
'catching',
'fudges',
'eye',
'the',
'portrait',
'said',
'he',
'will',
'be',
'here',
'in',
'a',
'moment',
'he',
'is',
'just',
'finishing',
'a',
'letter',
'to',
'dumbledore',
'i',
'wish',
'him',
'luck',
'said',
'fudge',
'sounding',

'bitter',
'for',
'the',
'first',
'time',
'i',
'have',
'been',
'writing',
'to',
'dumbledore',
'twice',
'a',
'day',
'for',
'the',
'past',
'fortnight',
'but',
'he',
'will',
'not',
'budge',
'if',
'he',
'would',
'just',
'been',
'prepared',
'to',
'persuade',
'the',
'boy',
'i',
'might',
'still',
'be',
'well',
'maybe',
'scrimgeour',
'will',
'have',
'more',
'success',
'fudge',
'subsided',
'into',
'what',
'was',
'clearly',
'an',
'aggrieved',
'silence',
'but',
'it',
'was',
'broken',
'almost',
'immediately',
'by',

'the',
'portrait',
'which',
'suddenly',
'spoke',
'in',
'its',
'crisp',
'official',
'voice',
'to',
'the',
'prime',
'minister',
'of',
'muggles',
'requesting',
'a',
'meeting',
'urgent',
'kindly',
'respond',
'immediately',
'rufus',
'scrimgeour',
'minister',
'of',
'magic',
'yes',
'yes',
'fine',
'said',
'the',
'prime',
'minister',
'distractedly',
'and',
'he',
'barely',
'flinched',
'as',
'the',
'flames',
'in',
'the',
'grate',
'turned',
'emerald',
'green',
'again',
'rose',
'up',
'and',
'revealed',
'a',
'second',
'spinning',
'wizard',
'in',
'their',

'heart',
'disgorging',
'him',
'moments',
'later',
'onto',
'the',
'antique',
'rug',
'fudge',
'got',
'to',
'his',
'feet',
'and',
'after',
'a',
'moments',
'hesitation',
'the',
'prime',
'minister',
'did',
'the',
'same',
'watching',
'the',
'new',
'arrival',
'straighten',
'up',
'dust',
'down',
'his',
'long',
'black',
'robes',
'and',
'look',
'around',
'the',
'prime',
'ministers',
'first',
'foolish',
'thought',
'was',
'that',
'rufus',
'scrimgeour',
'looked',
'rather',
'like',
'an',
'old',
'lion',
'there',
'were',
'streaks',
'of',

'gray',
'in',
'his',
'mane',
'of',
'tawny',
'hair',
'and',
'his',
'bushy',
'eyebrows',
'he',
'had',
'keen',
'yellowish',
'eyes',
'behind',
'a',
'pair',
'of',
'wirerimmed',
'spectacles',
'and',
'a',
'certain',
'rangy',
'loping',
'grace',
'even',
'though',
'he',
'walked',
'with',
'a',
'slight',
'limp',
'there',
'was',
'an',
'immediate',
'impression',
'of',
'shrewdness',
'and',
'toughness',
'the',
'prime',
'minister',
'thought',
'he',
'understood',
'why',
'the',
'wizarding',
'community',
'preferred',
'scrimgeour',
'to',
'fudge',
'as',

'a',
'leader',
'in',
'these',
'dangerous',
'times',
'how',
'do',
'you',
'do',
'said',
'the',
'prime',
'minister',
'politely',
'holding',
'out',
'his',
'hand',
'scrimgeour',
'grasped',
'it',
'briefly',
'his',
'eyes',
'scanning',
'the',
'room',
'then',
'pulled',
'out',
'a',
'wand',
'from',
'under',
'his',
'robes',
'fudge',
'told',
'you',
'everything',
'he',
'asked',
'striding',
'over',
'to',
'the',
'door',
'and',
'tapping',
'the',
'keyhole',
'with',
'his',
'wand',
'the',
'prime',
'minister',
'heard',
'the',

'lock',
'click',
'er',
'yes',
'said',
'the',
'prime',
'minister',
'and',
'if',
'you',
'do',
'not',
'mind',
'i',
'would',
'rather',
'that',
'door',
'remained',
'unlocked',
'i',
'would',
'rather',
'not',
'be',
'interrupted',
'said',
'scrimgeour',
'shortly',
'or',
'watched',
'he',
'added',
'pointing',
'his',
'wand',
'at',
'the',
'windows',
'so',
'that',
'the',
'curtains',
'swept',
'across',
'them',
'right',
'well',
'i',
'am',
'a',
'busy',
'man',
'so',
'let',
'us',
'get',
'down',
'to',

```
         'business',
         'first',
         'of',
         'all',
         'we',
         'need',
         'to',
         'discuss',
         'your',
         'security',
         'the',
         'prime',
         'minister',
         'drew',
         'himself',
         'up',
         'to',
         'his',
         'fullest',
         'height',
         'and',
         'replied',
         'i',
         'am',
         'perfectly',
         'happy',
         'with',
         'the',
         'security',
         'i',
         'have',
         'already',
         'got',
         'thank',
         'you',
         'very',
         'well',
         'we',
         'are',
         'not',
         ...]
```

In [59]: `reducer(words)`

```
Out[59]:  {'a': 89,
           'about': 3,
           'across': 3,
           'added': 1,
           'addled': 1,
           'admonitory': 1,
           'advisory': 1,
           'after': 4,
           'again': 2,
           'against': 1,
           'aggrieved': 1,
           'ago': 1,
           'ahead': 1,
           'air': 2,
           'all': 9,
           'alley': 1,
           'allow': 1,
           'almost': 3,
           'alone': 4,
           'already': 6,
           'am': 6,
           'amelia': 2,
           'an': 11,
           'and': 52,
           'anger': 1,
           'animal': 1,
           'another': 3,
           'antique': 1,
           'anxiously': 1,
           'any': 2,
           'anymore': 2,
           'anyone': 1,
           'anything': 2,
           'apart': 2,
           'appeared': 1,
           'are': 12,
           'arm': 3,
           'armchair': 2,
           'around': 6,
           'arrival': 1,
           'as': 21,
           'aside': 1,
           'asked': 5,
           'assigned': 1,
           'at': 16,
           'attacking': 1,
           'attempt': 1,
           'attempted': 2,
           'auror': 2,
           'away': 2,
           'azkaban': 1,
           'back': 7,
           'backyard': 1,
           'baffled': 1,
           'bang': 1,
           'bank': 4,
           'banks': 1,
           'barely': 3,
           'be': 11,
           'bearings': 1,
```

```
'because': 2,
'been': 7,
'before': 1,
'behind': 6,
'believe': 1,
'bella': 9,
'bellatrix': 2,
'beneath': 1,
'best': 1,
'betrayal': 1,
'between': 3,
'bit': 1,
'bitter': 1,
'black': 5,
'blind': 1,
'blonde': 1,
'blow': 1,
'boarded': 1,
'bones': 2,
'books': 2,
'bound': 1,
'bowler': 1,
'boy': 1,
'brains': 1,
'brave': 1,
'breakdown': 1,
'breath': 1,
'breathed': 1,
'breathing': 1,
'breeding': 1,
'breeze': 1,
'brick': 2,
'briefly': 1,
'bright': 1,
'bring': 1,
'broken': 3,
'brought': 1,
'brown': 1,
'budge': 1,
'burned': 1,
'burst': 1,
'bushy': 1,
'business': 1,
'busy': 3,
'but': 19,
'by': 5,
'can': 5,
'candlefilled': 1,
'cannot': 3,
'capacity': 1,
'carried': 1,
'case': 2,
'cast': 2,
'catching': 3,
'caught': 3,
'causing': 1,
'cautiously': 1,
'ceiling': 1,
'cell': 1,
'center': 1,
```

```
'certain': 1,
'chair': 1,
'check': 1,
'chilly': 1,
'chimney': 2,
'chorley': 1,
'cissy': 5,
'clasped': 1,
'cleared': 1,
'clearly': 2,
'click': 1,
'cloak': 4,
'closed': 1,
'cobbled': 1,
'cobbles': 1,
'coldly': 1,
'come': 1,
'community': 2,
'completely': 1,
'consented': 1,
'contempt': 1,
'continued': 1,
'continues': 1,
'corner': 2,
'could': 2,
'counting': 1,
'countryside': 1,
'course': 4,
'covered': 1,
'crack': 1,
'creatures': 2,
'crisp': 1,
'crouching': 1,
'cry': 1,
'curling': 1,
'curly': 1,
'curse': 2,
'cursing': 1,
'curtains': 3,
'curtly': 1,
'cut': 1,
'dangerous': 2,
'dark': 5,
'darkness': 3,
'darting': 1,
'date': 1,
'dawning': 1,
'day': 2,
'days': 1,
'dead': 1,
'dear': 1,
'decide': 1,
'decision': 1,
'declared': 1,
'deep': 1,
'deeper': 1,
'dementors': 2,
'department': 1,
'deserted': 2,
'despair': 1,
```

'despite': 1,
'developments': 1,
'did': 8,
'digging': 1,
'dilapidated': 1,
'dim': 2,
'directly': 1,
'dirty': 2,
'discuss': 1,
'disgorging': 1,
'dismissively': 1,
'distance': 1,
'distractedly': 1,
'disused': 1,
'diverted': 1,
'do': 10,
'does': 1,
'door': 7,
'down': 7,
'downstairs': 1,
'drain': 1,
'drew': 2,
'drifted': 1,
'drink': 1,
'drowned': 1,
'duck': 1,
'dull': 1,
'dumbledore': 2,
'dunghill': 1,
'dust': 1,
'each': 1,
'ear': 1,
'easy': 1,
'echoed': 1,
'edge': 1,
'efficient': 1,
'effort': 1,
'emerald': 1,
'emmeline': 1,
'end': 2,
'enforcement': 1,
'enough': 1,
'entertaining': 1,
'er': 2,
'even': 1,
'evening': 1,
'events': 1,
'ever': 1,
'everything': 1,
'evidence': 1,
'exactly': 1,
'examining': 1,
'excellent': 1,
'exchanged': 1,
'eye': 1,
'eyebrows': 1,
'eyes': 6,
'face': 2,
'faced': 1,
'fact': 1,

```
'faint': 2,
'fair': 1,
'fallen': 1,
'far': 1,
'feel': 1,
'feeling': 1,
'feet': 1,
'fell': 2,
'felt': 1,
'few': 2,
'field': 1,
'fight': 1,
'figure': 4,
'finally': 1,
'fine': 1,
'finger': 1,
'finishing': 1,
'fire': 2,
'fireplace': 1,
'first': 5,
'fishandchip': 1,
'fixed': 1,
'flames': 1,
'flash': 3,
'flat': 1,
'flew': 1,
'flicker': 1,
'flinched': 1,
'followed': 4,
'foolish': 1,
'foot': 1,
'footsteps': 1,
'for': 16,
'fortnight': 2,
'fought': 1,
'fox': 6,
'from': 15,
'froze': 1,
'fudge': 15,
'fudges': 1,
'fullest': 1,
'further': 1,
'gained': 1,
'gap': 1,
'gave': 1,
'gaze': 1,
'gazed': 1,
'gestured': 1,
'get': 1,
'gets': 3,
'getting': 1,
'giant': 1,
'gifted': 1,
'gleamed': 1,
'glimmered': 1,
'go': 4,
'going': 1,
'got': 4,
'grace': 1,
'grasped': 1,
```

```
'grass': 2,
'grate': 1,
'gray': 1,
'green': 4,
'ground': 1,
'grouped': 1,
'guard': 1,
'had': 18,
'hair': 3,
'hand': 3,
'hands': 1,
'happened': 1,
'happier': 1,
'happiness': 1,
'happy': 2,
'harsh': 1,
'has': 6,
'hat': 1,
'have': 11,
'he': 38,
'head': 1,
'healers': 1,
'hear': 2,
'heard': 2,
'heart': 1,
'heavens': 1,
'heavily': 1,
'height': 1,
'her': 30,
'herbert': 1,
'here': 9,
'hesitation': 1,
'hewhomustnotbenamed': 2,
'hidden': 1,
'hiding': 1,
'highly': 2,
'him': 15,
'himself': 2,
'his': 20,
'hold': 1,
'holding': 2,
'honestly': 1,
'hood': 4,
'hooded': 2,
'hope': 1,
'hopefully': 1,
'hopelessly': 1,
'hopelessness': 1,
'horror': 1,
'hospital': 1,
'hotly': 1,
'house': 2,
'houses': 3,
'hover': 1,
'how': 1,
'hung': 1,
'hurried': 1,
'hurrying': 1,
'hysteria': 1,
'i': 37,
```

'idea': 1,
'identical': 1,
'if': 6,
'immediate': 1,
'immediately': 2,
'immense': 1,
'imperius': 2,
'impersonating': 1,
'impression': 1,
'in': 39,
'incredulous': 1,
'indeed': 1,
'indignation': 1,
'inhabited': 1,
'injuries': 1,
'inside': 1,
'interrupted': 1,
'into': 10,
'introduce': 1,
'invisible': 1,
'invitation': 1,
'is': 25,
'it': 22,
'its': 5,
'jaw': 1,
'joined': 1,
'junior': 1,
'just': 8,
'keen': 1,
'keep': 1,
'keeping': 1,
'keyhole': 1,
'killed': 1,
'killing': 1,
'kind': 2,
'kindly': 2,
'kingsley': 1,
'knife': 1,
'knocked': 1,
'know': 1,
'known': 1,
'labyrinth': 1,
'lamely': 1,
'lamp': 1,
'lap': 1,
'last': 3,
'later': 1,
'laughed': 1,
'law': 2,
'leader': 1,
'leapt': 1,
'least': 1,
'leather': 1,
'leave': 1,
'left': 1,
'let': 3,
'letter': 1,
'lidded': 1,
'life': 1,
'light': 8,

```
'like': 3,
'limp': 1,
'line': 1,
'lion': 1,
'listen': 2,
'listened': 1,
'listening': 2,
'little': 2,
'lived': 1,
'lives': 1,
'lock': 1,
'locked': 1,
'long': 6,
'look': 3,
'looked': 5,
'looking': 2,
'lookout': 1,
'loping': 1,
'lord': 2,
'lords': 1,
'lost': 1,
'lot': 1,
'louder': 1,
'lowered': 1,
'luck': 1,
'made': 2,
'magic': 5,
'magical': 2,
'maladies': 1,
'man': 6,
'manage': 1,
'mane': 1,
'many': 1,
'materialized': 1,
'matter': 1,
'may': 2,
'maybe': 3,
'me': 5,
'meeting': 1,
'merely': 3,
'middleaged': 1,
'might': 1,
'miles': 1,
'mill': 2,
'mind': 1,
'minister': 31,
'ministers': 3,
'mist': 2,
'mistaken': 1,
'mocking': 1,
'moment': 4,
'momentarily': 3,
'moments': 3,
'more': 2,
'most': 1,
'mouth': 1,
'moved': 2,
'movement': 1,
'moving': 1,
'much': 1,
```

```
'muggle': 2,
'muggles': 2,
'mungos': 1,
'murder': 1,
'murdered': 1,
'must': 3,
'my': 6,
'mysterious': 1,
'named': 2,
'narcissa': 14,
'narrow': 1,
'nasty': 1,
'nearest': 1,
'need': 1,
'neglect': 1,
'never': 1,
'new': 3,
'newspapers': 2,
'night': 1,
'no': 3,
'nose': 1,
'not': 25,
'note': 1,
'nothing': 2,
'now': 7,
'of': 62,
'off': 2,
'office': 3,
'official': 1,
'oh': 1,
'old': 5,
'ominous': 1,
'on': 10,
'once': 2,
'one': 3,
'only': 1,
'onto': 1,
'open': 1,
'opened': 1,
'opening': 1,
'opposite': 2,
'or': 3,
'order': 1,
'other': 6,
'others': 1,
'our': 2,
'out': 6,
'outer': 1,
'over': 6,
'overgrown': 1,
'own': 1,
'padded': 1,
'pair': 2,
'pale': 1,
'panted': 1,
'panting': 1,
'papers': 1,
'parted': 1,
'pass': 1,
'passed': 2,
```

```
'past': 1,
'patches': 1,
'paused': 1,
'people': 3,
'perfectly': 1,
'performed': 1,
'perhaps': 1,
'person': 2,
'personally': 1,
'persuade': 1,
'phenomenon': 1,
'place': 3,
'placed': 1,
'plan': 1,
'pleasant': 1,
'pocket': 1,
'point': 1,
'pointed': 1,
'pointing': 1,
'police': 1,
'politely': 1,
'pool': 1,
'poor': 1,
'poorly': 1,
'pop': 2,
'portrait': 3,
'position': 1,
'posted': 1,
'powder': 1,
'preferred': 1,
'prepared': 1,
'pressed': 1,
'pretend': 1,
'prey': 1,
'prime': 30,
'prison': 1,
'prisoners': 1,
'probably': 1,
'problem': 1,
'protection': 1,
'public': 1,
'publicity': 1,
'pulled': 1,
'pursuer': 2,
'put': 3,
'quacking': 1,
'quarry': 1,
'quick': 1,
'quietly': 1,
'quill': 1,
'quite': 1,
'railings': 2,
'rangy': 1,
'rather': 6,
'reached': 1,
'reacted': 1,
'real': 1,
'really': 2,
'reared': 1,
'recent': 1,
```

```
'relic': 1,
'remained': 1,
'remove': 1,
'replied': 2,
'requesting': 1,
'resignation': 1,
'respond': 1,
'responsibility': 1,
'rest': 2,
'revealed': 1,
'revealing': 1,
'rickety': 1,
'rid': 1,
'right': 4,
'river': 4,
'road': 2,
'robes': 2,
'room': 4,
'rose': 1,
'rows': 2,
'rubbing': 1,
'rubbishstrewn': 1,
'rufus': 2,
'rug': 1,
'rummaging': 1,
'running': 1,
'rushed': 1,
'rustling': 1,
'rusty': 1,
'sacked': 1,
'said': 30,
'sake': 1,
'sallow': 1,
'same': 1,
'sank': 1,
'sat': 1,
'saw': 1,
'say': 2,
'scanning': 1,
'scrambling': 1,
'scrawny': 1,
'screaming': 1,
'scrimgeour': 16,
'second': 5,
'seconds': 1,
'secretary': 1,
'security': 2,
'see': 2,
'seemed': 5,
'seen': 1,
'seized': 1,
'send': 1,
'sense': 1,
'sent': 1,
'sentence': 1,
'separated': 1,
'set': 2,
'settling': 1,
'severus': 1,
'shacklebolt': 2,
```

```
'shacklebolts': 1,
'shadowy': 1,
'shall': 2,
'she': 20,
'shine': 1,
'shortly': 1,
'shrewdness': 1,
'shrugged': 1,
'shrunkenlooking': 1,
'side': 3,
'sighed': 1,
'sign': 1,
'silence': 1,
'silver': 1,
'sister': 4,
'sisters': 2,
'sitting': 2,
'slight': 1,
'slightly': 2,
'slim': 1,
'slipped': 1,
'sliver': 1,
'slowly': 2,
'slunk': 1,
'smell': 1,
'smile': 5,
'snap': 1,
'snape': 4,
'snarled': 1,
'so': 9,
'society': 1,
'sofa': 2,
'some': 2,
'something': 1,
'sorry': 1,
'sort': 1,
'sound': 1,
'sounding': 1,
'speak': 3,
'spectacles': 1,
'spinners': 2,
'spinning': 2,
'spoke': 1,
'spot': 1,
'spreading': 1,
'st': 1,
'stand': 1,
'staring': 1,
'startled': 1,
'stay': 1,
'stepped': 2,
'still': 4,
'stillhooded': 1,
'stood': 4,
'stopped': 1,
'straighten': 1,
'strained': 1,
'strange': 1,
'strangle': 1,
'streaks': 1,
```

```
'streaming': 2,
'street': 3,
'streetlamps': 1,
'strides': 1,
'striding': 1,
'strong': 1,
'subsided': 1,
'succeeding': 1,
'success': 1,
'successor': 1,
'suddenly': 1,
'suggesting': 1,
'suppress': 1,
'surely': 2,
'surprise': 1,
'swarming': 1,
'swept': 1,
'swinging': 1,
'swooping': 1,
'table': 1,
'take': 2,
'tall': 1,
'tapping': 1,
'tawny': 1,
'team': 1,
'tell': 1,
'term': 1,
'thank': 1,
'that': 29,
'the': 157,
'their': 3,
'them': 9,
'then': 6,
'there': 11,
'these': 1,
'they': 12,
'thin': 2,
'think': 3,
'this': 9,
'though': 4,
'thought': 6,
'threadbare': 1,
'threateningly': 1,
'three': 2,
'threw': 2,
'throat': 1,
'through': 5,
'time': 4,
'times': 1,
'tiny': 1,
'to': 45,
'toe': 1,
'together': 2,
'told': 2,
'tonight': 1,
'too': 3,
'toothache': 1,
'top': 1,
'toughness': 1,
'toward': 2,
```

```
'towering': 1,
'towns': 1,
'trained': 1,
'trembling': 1,
'trouble': 1,
'trust': 1,
'trusts': 1,
'turned': 5,
'twice': 2,
'two': 3,
'ugly': 1,
'under': 5,
'undergrowth': 1,
'understood': 1,
'unintelligible': 1,
'united': 1,
'unlocked': 1,
'unsuccessful': 1,
'until': 1,
'up': 11,
'upon': 3,
'urgent': 2,
'us': 2,
'usually': 1,
'vance': 1,
'vanished': 1,
'vermin': 1,
'very': 7,
'voice': 4,
'voters': 1,
'wait': 4,
'waiting': 1,
'walked': 1,
'wall': 1,
'walls': 1,
'wand': 6,
'wary': 1,
'was': 33,
'watched': 1,
'watching': 1,
'water': 1,
'we': 14,
'weakkneed': 1,
'weakly': 1,
'wearily': 1,
'wearing': 1,
'well': 9,
'were': 7,
'what': 6,
'where': 2,
'which': 5,
'while': 1,
'whisper': 2,
'white': 1,
'who': 8,
'whole': 2,
'why': 1,
'wider': 1,
'wig': 1,
'will': 8,
```

```
           'windows': 4,
           'wirerimmed': 1,
           'wiser': 1,
           'wish': 1,
           'witch': 1,
           'with': 21,
           'without': 2,
           'wizard': 2,
           'wizarding': 2,
           'wizards': 2,
           'woman': 4,
           'womans': 1,
           'women': 1,
           'words': 2,
           'work': 2,
           'works': 1,
           'wormtails': 1,
           'would': 7,
           'wound': 1,
           'wrappings': 1,
           'wrenched': 1,
           'writing': 1,
           'yellowish': 1,
           'yelp': 1,
           'yes': 5,
           'you': 26,
           'your': 7}
```

In [60]:
```python
def map_non_english(text,english_words):
    """
    This functions returns tokenized non english words list from the given text
    """
    words = text.split(' ')
    non_english_words = []
    for i in words:
        if i not in english_words:
            non_english_words.append(i)
        else:
            continue
    return non_english_words
```

In [64]:
```python
non_english_words = map_non_english(parsed_text2,english_words)
non_english_words
```

```
Out[64]: ['harrys',
          'dursleys',
          'harrys',
          'hedwig',
          'threequarters',
          'hogwarts',
          'hogwarts',
          'hagrid',
          'diagon',
          'muggles',
          'harrys',
          'threequarters',
          'hogwarts',
          'rons',
          'hogwarts',
          'wroughtiron',
          'threequarters',
          'roundfaced',
          'hedwig',
          'redhaired',
          'cmere',
          'harrys',
          'harrys',
          'redhaired',
          'geroff',
          'aaah',
          'ickle',
          'somefink',
          'nosie',
          'hogwarts',
          'ickle',
          'ronniekins',
          'blackhaired',
          'youknowwho',
          'hogwarts',
          'jordans',
          'weasley',
          'harrys',
          'youknowwho',
          'weasleys',
          'wizarding',
          'diagon',
          'muggles',
          'hogwarts',
          'quidditch',
          'percys']
```

```
In [65]: reducer(non_english_words)
```

```
Out[65]:    {'aaah': 1,
            'blackhaired': 1,
            'cmere': 1,
            'diagon': 2,
            'dursleys': 1,
            'geroff': 1,
            'hagrid': 1,
            'harrys': 6,
            'hedwig': 2,
            'hogwarts': 7,
            'ickle': 2,
            'jordans': 1,
            'muggles': 2,
            'nosie': 1,
            'percys': 1,
            'quidditch': 1,
            'redhaired': 2,
            'ronniekins': 1,
            'rons': 1,
            'roundfaced': 1,
            'somefink': 1,
            'threequarters': 3,
            'weasley': 1,
            'weasleys': 1,
            'wizarding': 1,
            'wroughtiron': 1,
            'youknowwho': 2}
```