11_Word_Frequency

August 14, 2017

In [57]:
```
import os
from string import punctuation
import pandas as pd
import nltk
from nltk.tokenize import RegexpTokenizer
from nltk.stem.wordnet import WordNetLemmatizer
```

In [2]:
```
lmtzr = WordNetLemmatizer()
tokenizer = RegexpTokenizer(r'\w+')
stopwords = nltk.corpus.stopwords.words('english')

def processLine(line):
    tokens = tokenizer.tokenize(line)
    output = []
    for t in tokens:
        t = lmtzr.lemmatize(t)
        if (len(t) <= 2) or (len(t) >= 15) or (t in stopwords) or (t.isdigit()):
            continue
        output.append(t)
    return ' '.join(output)
```

In [92]:
```
manufacture = ['MENTOR', 'ALLERGAN', 'IDEAL', 'SIENTRA', 'MCGHAN', 'INAMED', 'SILIMED
fill_type = ['SALINE', 'SILICONE', 'GEL', 'COHESIVE']
surface_type = ['SMOOTH', 'TEXTURED', 'BIOCELL']
implantation_indication = ['AUGMENTATION', 'RECONSTRUCTION', 'COSMETIC']
ALCL = ['ALCL', 'ANAPLASTIC LARGE CELL LYMPHOMA', 'LYMPHOMA', 'T-CELL LYMPHOMA', 'B-C
side = ['LEFT', 'RIGHT']
symptom = ['breast pain', 'breast swelling', 'breast cyst', 'breast calcification', '
           'lymph node enlargement', 'firmness of breast''hematoma', 'mass', 'lump',
           'infection', 'abscess', 'leukopenia', 'nodules', 'skin discoloration', 'ski
           'effusion', 'fluid']
word_list = [manufacture, fill_type, surface_type, implantation_indication, ALCL, side

MDR_KEY_LIST = []
with open('Data/No time filter/FWM_Key_no_time_filter.txt') as fh:
    lines = fh.readlines()
    for line in lines:
```

```
                MDR_KEY_LIST.append(line.rstrip('\n'))
        with open('Data/No time filter/FTR_Key_no_time_filter.txt') as fh:
            lines = fh.readlines()
            for line in lines:
                MDR_KEY_LIST.append(line.rstrip('\n'))
```

```
In [4]: processed = []
        for l in word_list:
            for w in l:
                p = processLine(w)
                print(w, ':', p)
```

```
MENTOR : mentor
ALLERGAN : allergan
IDEAL : ideal
SIENTRA : sientra
MCGHAN : mcghan
INAMED : inamed
SILIMED : silimed
NAGOR : nagor
SALINE : saline
SILICONE : silicone
COHESIVE : cohesive
SMOOTH : smooth
TEXTURED : textured
BIOCELL : biocell
AUGMENTATION : augmentation
RECONSTRUCTION : reconstruction
COSMETIC : cosmetic
ANAPLASTIC LARGE CELL LYMPHOMA : anaplastic large cell lymphoma
LYMPHOMA : lymphoma
T-CELL LYMPHOMA : cell lymphoma
B-CELL LYMPHOMA : cell lymphoma
left : left
right : right
breast pain : breast pain
breast swelling : breast swelling
breast cyst : breast cyst
breast calcification : breast calcification
capsular contracture : capsular contracture
lymph node enlargement : lymph node enlargement
firmness of breasthematoma : firmness breasthematoma
mass : mass
lump : lump
rupture : rupture
deflated : deflated
infection : infection
abscess : abscess
```

```
leukopenia : leukopenia
nodules : nodule
skin discoloration : skin discoloration
skin lesion : skin lesion
seroma : seroma
effusion : effusion
fluid : fluid
```

```
In [26]: def filtering(s):
             if (s == 'FWM') | (s == 'FTR'):
                 return True
             else:
                 return False
```

```
In [27]: FOI_DEV_LIST = ['', 'Add', 'Change', 'thru1997']
         for i in range(1998, 2017):
             FOI_DEV_LIST.append(str(i))

         df_list = []
         for s in FOI_DEV_LIST:
             df = pd.read_csv('foidev/foidev'+s+'.txt', sep='|', header=0, encoding='ISO-8859-1
             df_list.append(df)
         df_BASELINE = pd.concat(df_list, axis=0)
         del df_list

         df_BASELINE['filter'] = df_BASELINE['DEVICE_REPORT_PRODUCT_CODE'].map(filtering)
         df_BI_DEV = df_BASELINE.loc[df_BASELINE['filter']==True, :]
```

```
b'Skipping line 3974: expected 28 fields, saw 29\n'
b'Skipping line 46727: expected 28 fields, saw 29\n'
b'Skipping line 7949: expected 28 fields, saw 29\n'
b'Skipping line 24283: expected 28 fields, saw 29\n'
b'Skipping line 54015: expected 45 fields, saw 47\n'
b'Skipping line 66558: expected 45 fields, saw 58\n'
b'Skipping line 121357: expected 45 fields, saw 59\nSkipping line 122019: expected 45 fields, s
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/IPython/core/inte
  interactivity=interactivity, compiler=compiler, result=result)
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/IPython/core/inte
  interactivity=interactivity, compiler=compiler, result=result)
b'Skipping line 16452: expected 45 fields, saw 46\n'
b'Skipping line 48741: expected 45 fields, saw 57\n'
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/IPython/core/inte
  interactivity=interactivity, compiler=compiler, result=result)
b'Skipping line 23599: expected 45 fields, saw 48\n'
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/IPython/core/inte
  interactivity=interactivity, compiler=compiler, result=result)
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/IPython/core/inte
```

```
  interactivity=interactivity, compiler=compiler, result=result)
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/IPython/core/inte
  interactivity=interactivity, compiler=compiler, result=result)
b'Skipping line 34672: expected 28 fields, saw 29\n'
b'Skipping line 117249: expected 28 fields, saw 29\n'
b'Skipping line 154198: expected 28 fields, saw 29\n'
b'Skipping line 211424: expected 28 fields, saw 29\n'
b'Skipping line 267765: expected 28 fields, saw 29\n'
b'Skipping line 397060: expected 28 fields, saw 29\n'
b'Skipping line 426436: expected 28 fields, saw 29\n'
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/IPython/core/inte
  interactivity=interactivity, compiler=compiler, result=result)
b'Skipping line 24816: expected 28 fields, saw 29\n'
b'Skipping line 159206: expected 28 fields, saw 29\n'
b'Skipping line 166194: expected 28 fields, saw 29\nSkipping line 187565: expected 28 fields, s
b'Skipping line 285177: expected 28 fields, saw 42\n'
b'Skipping line 336730: expected 28 fields, saw 42\n'
b'Skipping line 516611: expected 28 fields, saw 29\n'
b'Skipping line 537510: expected 28 fields, saw 42\n'
b'Skipping line 633833: expected 28 fields, saw 29\n'
b'Skipping line 670754: expected 28 fields, saw 29\n'
b'Skipping line 725216: expected 28 fields, saw 29\n'
b'Skipping line 844969: expected 28 fields, saw 29\n'
b'Skipping line 99668: expected 28 fields, saw 40\n'
b'Skipping line 397258: expected 28 fields, saw 29\n'
b'Skipping line 535712: expected 28 fields, saw 29\n'
b'Skipping line 648332: expected 28 fields, saw 29\n'
b'Skipping line 707376: expected 28 fields, saw 29\n'
b'Skipping line 839606: expected 28 fields, saw 29\n'
b'Skipping line 10097: expected 28 fields, saw 29\nSkipping line 12356: expected 28 fields, saw
b'Skipping line 88785: expected 28 fields, saw 29\n'
b'Skipping line 202635: expected 28 fields, saw 29\n'
b'Skipping line 230410: expected 28 fields, saw 29\nSkipping line 232766: expected 28 fields, s
b'Skipping line 263700: expected 28 fields, saw 29\nSkipping line 289010: expected 28 fields, s
b'Skipping line 334936: expected 28 fields, saw 29\n'
b'Skipping line 443377: expected 28 fields, saw 29\n'
b'Skipping line 570165: expected 28 fields, saw 29\n'
b'Skipping line 606191: expected 28 fields, saw 29\nSkipping line 620211: expected 28 fields, s
b'Skipping line 694779: expected 28 fields, saw 29\n'
b'Skipping line 729113: expected 28 fields, saw 29\nSkipping line 733739: expected 28 fields, s
b'Skipping line 775832: expected 28 fields, saw 29\n'
b'Skipping line 786838: expected 28 fields, saw 29\n'
```

```
In [35]: print(df_BI_DEV.shape[0])
         # del df_BI_DEV['filter']
         for n in df_BI_DEV.columns.values:
             if 'BASELINE' in n:
```

```
                del df_BI_DEV[n]
        df_BI_DEV.to_csv('WF/DEV_BI_FULL_TABLE.txt', header=True, index=False, sep='|')
```

```
29031
```

```
In [42]: print('Manufacturer word frequecy:\n')
         for m in manufacture:
             counter = 0
             fh = open('WF/manufacturer/'+m+'.txt', 'w')
             for index, value in df_BI_DEV.iterrows():
                 l = str(value['MANUFACTURER_D_NAME']).split(' ')
                 if m in l:
                     fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                     counter += 1
             print(m, ':', counter)
             fh.close()
```

```
Manufacturer word frequecy:

MENTOR : 3439
ALLERGAN : 2100
IDEAL : 24
SIENTRA : 23
MCGHAN : 2120
INAMED : 189
SILIMED : 7
NAGOR : 4
```

```
In [62]: FOI_TEXT_LIST = ['', 'Add', 'Change', 'thru1995']
         for i in range(1996, 2017):
             FOI_TEXT_LIST.append(str(i))

         df_BI = pd.DataFrame(df_BI_DEV['MDR_REPORT_KEY'])
         df_BI_TEXT = pd.DataFrame()
         for s in FOI_TEXT_LIST:
             df = pd.read_csv('foitext/foitext'+s+'.txt', sep='|', header=0, encoding='ISO-8859
             temp = df.merge(df_BI, on=['MDR_REPORT_KEY'], how='inner')
             df_BI_TEXT = pd.concat([df_BI_TEXT, temp], axis=0)
             del temp

         print(df_BI_TEXT.shape[0])
         df_BI_TEXT.head()
```

```
b'Skipping line 12326: expected 6 fields, saw 7\nSkipping line 41095: expected 6 fields, saw 7'
b'Skipping line 136392: expected 6 fields, saw 7\nSkipping line 175249: expected 6 fields, saw
b'Skipping line 278732: expected 6 fields, saw 7\nSkipping line 315087: expected 6 fields, saw
b'Skipping line 469829: expected 6 fields, saw 7\nSkipping line 471072: expected 6 fields, saw
```

```
b'Skipping line 548966: expected 6 fields, saw 8\nSkipping line 549426: expected 6 fields, saw
b'Skipping line 702080: expected 6 fields, saw 7\nSkipping line 768084: expected 6 fields, saw
b'Skipping line 889388: expected 6 fields, saw 8\n'
b'Skipping line 939394: expected 6 fields, saw 7\nSkipping line 953223: expected 6 fields, saw
b'Skipping line 1072605: expected 6 fields, saw 7\nSkipping line 1115860: expected 6 fields, sa
b'Skipping line 1909: expected 6 fields, saw 7\nSkipping line 106185: expected 6 fields, saw 7\
b'Skipping line 4144: expected 6 fields, saw 7\nSkipping line 18057: expected 6 fields, saw 7\n
b'Skipping line 186742: expected 6 fields, saw 7\nSkipping line 188236: expected 6 fields, saw
b'Skipping line 262909: expected 6 fields, saw 7\nSkipping line 325993: expected 6 fields, saw
b'Skipping line 16266: expected 6 fields, saw 28\nSkipping line 18280: expected 6 fields, saw 7
b'Skipping line 5693: expected 6 fields, saw 7\nSkipping line 18817: expected 6 fields, saw 7\n
b'Skipping line 3498: expected 6 fields, saw 7\nSkipping line 5044: expected 6 fields, saw 7\nS
b'Skipping line 140608: expected 6 fields, saw 8\nSkipping line 153010: expected 6 fields, saw
b'Skipping line 604861: expected 6 fields, saw 7\n'
b'Skipping line 279892: expected 6 fields, saw 10\n'
b'Skipping line 513515: expected 6 fields, saw 10\n'
b'Skipping line 70946: expected 6 fields, saw 7\nSkipping line 104363: expected 6 fields, saw 8
b'Skipping line 216807: expected 6 fields, saw 8\n'
b'Skipping line 372694: expected 6 fields, saw 8\nSkipping line 385546: expected 6 fields, saw
b'Skipping line 502732: expected 6 fields, saw 7\nSkipping line 502744: expected 6 fields, saw
b'Skipping line 614747: expected 6 fields, saw 7\n'
b'Skipping line 655835: expected 6 fields, saw 8\n'
b'Skipping line 801296: expected 6 fields, saw 7\nSkipping line 842014: expected 6 fields, saw
b'Skipping line 958137: expected 6 fields, saw 7\n'
b'Skipping line 1118369: expected 6 fields, saw 9\n'
b'Skipping line 1246859: expected 6 fields, saw 7\n'
b'Skipping line 236429: expected 6 fields, saw 7\n'
b'Skipping line 459280: expected 6 fields, saw 14\n'
b'Skipping line 1189476: expected 6 fields, saw 7\n'
b'Skipping line 1417218: expected 6 fields, saw 8\n'
b'Skipping line 1443165: expected 6 fields, saw 7\n'
b'Skipping line 58566: expected 6 fields, saw 18\nSkipping line 61716: expected 6 fields, saw 8
b'Skipping line 166628: expected 6 fields, saw 27\nSkipping line 173673: expected 6 fields, sau
b'Skipping line 335754: expected 6 fields, saw 9\nSkipping line 370578: expected 6 fields, saw
b'Skipping line 477356: expected 6 fields, saw 8\nSkipping line 508812: expected 6 fields, saw
b'Skipping line 577703: expected 6 fields, saw 7\nSkipping line 609991: expected 6 fields, saw
b'Skipping line 697385: expected 6 fields, saw 8\nSkipping line 715386: expected 6 fields, saw
b'Skipping line 791567: expected 6 fields, saw 8\nSkipping line 805014: expected 6 fields, saw
b'Skipping line 979034: expected 6 fields, saw 7\nSkipping line 1045512: expected 6 fields, sau
b'Skipping line 1082683: expected 6 fields, saw 9\nSkipping line 1114444: expected 6 fields, sa
b'Skipping line 1242335: expected 6 fields, saw 7\nSkipping line 1278243: expected 6 fields, sa
b'Skipping line 1350712: expected 6 fields, saw 7\nSkipping line 1413257: expected 6 fields, sa
b'Skipping line 1454198: expected 6 fields, saw 9\nSkipping line 1561372: expected 6 fields, sa
b'Skipping line 1582569: expected 6 fields, saw 7\nSkipping line 1632552: expected 6 fields, sa
b'Skipping line 1792767: expected 6 fields, saw 12\nSkipping line 1796204: expected 6 fields, s
b'Skipping line 1835849: expected 6 fields, saw 9\nSkipping line 1841427: expected 6 fields, sa
b'Skipping line 12054: expected 6 fields, saw 8\nSkipping line 12414: expected 6 fields, saw 7\
b'Skipping line 145016: expected 6 fields, saw 7\nSkipping line 145918: expected 6 fields, saw
```

```
b'Skipping line 266986: expected 6 fields, saw 8\nSkipping line 280145: expected 6 fields, saw
b'Skipping line 400137: expected 6 fields, saw 8\nSkipping line 404381: expected 6 fields, saw
b'Skipping line 534823: expected 6 fields, saw 7\nSkipping line 571038: expected 6 fields, saw
b'Skipping line 655663: expected 6 fields, saw 7\nSkipping line 659388: expected 6 fields, saw
b'Skipping line 791063: expected 6 fields, saw 7\nSkipping line 792799: expected 6 fields, saw
b'Skipping line 918365: expected 6 fields, saw 11\nSkipping line 931557: expected 6 fields, sa
b'Skipping line 1091634: expected 6 fields, saw 13\nSkipping line 1093872: expected 6 fields, s
b'Skipping line 1181580: expected 6 fields, saw 8\nSkipping line 1202211: expected 6 fields, sa
b'Skipping line 1320743: expected 6 fields, saw 7\nSkipping line 1355491: expected 6 fields, sa
b'Skipping line 1450797: expected 6 fields, saw 9\nSkipping line 1456625: expected 6 fields, sa
b'Skipping line 1577001: expected 6 fields, saw 7\nSkipping line 1588271: expected 6 fields, sa
b'Skipping line 1706169: expected 6 fields, saw 8\nSkipping line 1716700: expected 6 fields, sa
b'Skipping line 1844159: expected 6 fields, saw 21\nSkipping line 1852158: expected 6 fields, s
b'Skipping line 2003027: expected 6 fields, saw 7\nSkipping line 2012099: expected 6 fields, sa
b'Skipping line 2098204: expected 6 fields, saw 7\nSkipping line 2143443: expected 6 fields, sa
b'Skipping line 2230573: expected 6 fields, saw 8\nSkipping line 2231276: expected 6 fields, sa
b'Skipping line 1427: expected 6 fields, saw 7\nSkipping line 35232: expected 6 fields, saw 8\n
b'Skipping line 215408: expected 6 fields, saw 7\n'
b'Skipping line 263879: expected 6 fields, saw 9\nSkipping line 315456: expected 6 fields, saw
b'Skipping line 411784: expected 6 fields, saw 7\n'
b'Skipping line 557447: expected 6 fields, saw 7\nSkipping line 583140: expected 6 fields, saw
b'Skipping line 701851: expected 6 fields, saw 7\nSkipping line 722897: expected 6 fields, saw
b'Skipping line 809305: expected 6 fields, saw 7\nSkipping line 886939: expected 6 fields, saw
b'Skipping line 918501: expected 6 fields, saw 7\nSkipping line 1042125: expected 6 fields, sa
b'Skipping line 1159746: expected 6 fields, saw 7\n'
b'Skipping line 1249889: expected 6 fields, saw 23\n'
b'Skipping line 1372331: expected 6 fields, saw 7\nSkipping line 1412006: expected 6 fields, sa
b'Skipping line 1451832: expected 6 fields, saw 9\nSkipping line 1455944: expected 6 fields, sa
b'Skipping line 1595489: expected 6 fields, saw 10\nSkipping line 1608909: expected 6 fields, s
b'Skipping line 1704719: expected 6 fields, saw 7\nSkipping line 1812728: expected 6 fields, sa
b'Skipping line 1858898: expected 6 fields, saw 19\nSkipping line 1876811: expected 6 fields, s
b'Skipping line 2081242: expected 6 fields, saw 7\n'


34208


Out[62]:    MDR_REPORT_KEY  MDR_TEXT_KEY TEXT_TYPE_CODE  PATIENT_SEQUENCE_NUMBER  \
         0         6730886      80620016              D                        1
         1         6730886      80620016              D                        1
         2         6730886      80620016              D                        1
         3         6730886      80620015              N                        1
         4         6730886      80620015              N                        1

            DATE_REPORT                                           FOI_TEXT
         0          NaN  HEALTHCARE PROFESSIONAL REPORTED A LEFT SIDE D...
         1          NaN  HEALTHCARE PROFESSIONAL REPORTED A LEFT SIDE D...
         2          NaN  HEALTHCARE PROFESSIONAL REPORTED A LEFT SIDE D...
```

```
3          NaN  DEVICE EVALUATION: VISUAL ANALYSIS OF THE RETU...
4          NaN  DEVICE EVALUATION: VISUAL ANALYSIS OF THE RETU...
```

In [63]:
```
df_BI_TEXT = df_BI_TEXT.drop_duplicates('MDR_TEXT_KEY')
print(df_BI_TEXT.shape[0])
df_BI_TEXT.to_csv('WF/TEXT_FULL_BI_LIST.txt', header=True, index=False, sep='|')
df_BI_TEXT.head()
```

```
27137
```

Out[63]:

	MDR_REPORT_KEY	MDR_TEXT_KEY	TEXT_TYPE_CODE	PATIENT_SEQUENCE_NUMBER	DATE_REPORT	FOI_TEXT
0	6730886	80620016	D	1	NaN	HEALTHCARE PROFESSIONAL REPORTED A LEFT SIDE D...
3	6730886	80620015	N	1	NaN	DEVICE EVALUATION: VISUAL ANALYSIS OF THE RETU...
6	6734192	80750975	N	1	NaN	THE DEVICE REMAINS IMPLANTED. A REVIEW OF THE ...
9	6734192	80750976	D	1	NaN	PATIENT REPORTED THAT THE LEFT SIDE "HAS COLLA...
12	6283766	66229409	D	1	NaN	DEFLATION RESULTING IN EXPLANTATION.

In [71]:
```
punctuations = '''!()[]{};:'"\,<>./?@#$%^&*_~'''

def processing_text(s):
    s = str(s)
    no_punct = ""
    for char in s:
        if char not in punctuations:
            no_punct = no_punct + char
    l = no_punct.split(' ')
    output = []
    for w in l:
        if (w in stopwords) or (w.isdigit()):
            continue
        output.append(w)
    return ' '.join(output)
```

In [72]:
```
df_BI_TEXT['FOI_TEXT'] = df_BI_TEXT['FOI_TEXT'].map(processing_text)
df_BI_TEXT['FOI_TEXT'].head()
```

Out[72]:
```
0    HEALTHCARE PROFESSIONAL REPORTED A LEFT SIDE D...
1    DEVICE EVALUATION VISUAL ANALYSIS OF THE RETUR...
2    THE DEVICE REMAINS IMPLANTED A REVIEW OF THE D...
3    PATIENT REPORTED THAT THE LEFT SIDE HAS COLLAP...
4                  DEFLATION RESULTING IN EXPLANTATION
Name: FOI_TEXT, dtype: object
```

```
In [93]: print('Fill type word frequecy:\n')
         for m in fill_type:
             counter = 0
             fh = open('WF/fill_type/'+m+'.txt', 'w')
             for index, value in df_BI_TEXT.iterrows():
                 l = str(value['FOI_TEXT']).split(' ')
                 if m in l:
                     fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                     counter += 1
             print(m, ':', counter)
             fh.close()
```

```
Fill type word frequecy:

SALINE : 3309
SILICONE : 8184
GEL : 5113
COHESIVE : 74
```

```
In [75]: print('Surface type word frequecy:\n')
         for m in surface_type:
             counter = 0
             fh = open('WF/surface_type/'+m+'.txt', 'w')
             for index, value in df_BI_TEXT.iterrows():
                 l = str(value['FOI_TEXT']).split(' ')
                 if m in l:
                     fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                     counter += 1
             print(m, ':', counter)
             fh.close()
```

```
Surface type word frequecy:

SMOOTH : 372
TEXTURED : 244
BIOCELL : 17
```

```
In [76]: print('Implantation indication word frequecy:\n')
         for m in implantation_indication:
             counter = 0
             fh = open('WF/implantation_indication/'+m+'.txt', 'w')
             for index, value in df_BI_TEXT.iterrows():
                 l = str(value['FOI_TEXT']).split(' ')
                 if m in l:
                     fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                     counter += 1
```

```
            print(m, ':', counter)
            fh.close()
```

```
Implantation indication word frequecy:

AUGMENTATION : 1795
RECONSTRUCTION : 981
COSMETIC : 226
```

```
In [84]: print('ALCL word frequecy:\n')
         for m in ALCL:
             counter = 0
             fh = open('WF/ALCL/'+m+'.txt', 'w')
             for index, value in df_BI_TEXT.iterrows():
                 if m in str(value['FOI_TEXT']):
                     fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                     counter += 1
             print(m, ':', counter)
             fh.close()
```

```
ALCL word frequecy:

ALCL : 875
ANAPLASTIC LARGE CELL LYMPHOMA : 471
LYMPHOMA : 1033
T-CELL LYMPHOMA : 73
B-CELL LYMPHOMA : 1
```

```
In [87]: print('Side word frequecy:\n')
         for m in side:
             counter = 0
             fh = open('WF/side/'+m+'.txt', 'w')
             for index, value in df_BI_TEXT.iterrows():
                 if m in str(value['FOI_TEXT']):
                     fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                     counter += 1
             print(m, ':', counter)
             fh.close()
```

```
Side word frequecy:

LEFT : 6543
RIGHT : 6609
```

```
In [89]: print('Symptom word frequecy:\n')
         for m in symptom:
```

```
            m = m.upper()
            counter = 0
            fh = open('WF/symptom/'+m+'.txt', 'w')
            for index, value in df_BI_TEXT.iterrows():
                if m in str(value['FOI_TEXT']):
                    fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                    counter += 1
            print(m, ':', counter)
            fh.close()
```

```
Symptom word frequecy:

BREAST PAIN : 706
BREAST SWELLING : 35
BREAST CYST : 5
BREAST CALCIFICATION : 3
CAPSULAR CONTRACTURE : 2979
LYMPH NODE ENLARGEMENT : 3
FIRMNESS OF BREASTHEMATOMA : 0
MASS : 594
LUMP : 674
RUPTURE : 6573
DEFLATED : 1125
INFECTION : 1890
ABSCESS : 45
LEUKOPENIA : 2
NODULES : 76
SKIN DISCOLORATION : 10
SKIN LESION : 17
SEROMA : 1125
EFFUSION : 45
FLUID : 637
```

```
In [95]: df_BI_DEV['BRAND_NAME'] = df_BI_DEV['BRAND_NAME'].map(processing_text)
         df_BI_DEV['BRAND_NAME'].head()
```

```
/Library/Frameworks/Python.framework/Versions/3.6/lib/python3.6/site-packages/ipykernel_launche
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: http://pandas.pydata.org/pandas-docs/stable/indexing.html
  """Entry point for launching an IPython kernel.
```

```
Out[95]: 887                                        BREAST IMPLANTS
         888                                        BREAST IMPLANTS
         1705    SALINESILICONE BREAST IMPLANT RIGHT AND LEFT U...
         2263                    STYLE SALINE FILLED BREAST IMPLANT
```

```
2327                    STYLE SALINE FILLED BREAST IMPLANT
Name: BRAND_NAME, dtype: object
```

```
In [101]: print('Surface type word frequecy (brand name):\n')
          for m in surface_type:
              counter = 0
              fh = open('WF/surface_type_BRAND_NAME/'+m+'.txt', 'w')
              for index, value in df_BI_DEV.iterrows():
                  if m in str(value['BRAND_NAME']):
                      fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                      counter += 1
              print(m, ':', counter)
              fh.close()
```

```
Surface type word frequecy (brand name):

SMOOTH : 808
TEXTURED : 406
BIOCELL : 53
```

```
In [102]: print('Fill type word frequecy (brand name):\n')
          for m in fill_type:
              counter = 0
              fh = open('WF/fill_type_BRAND_NAME/'+m+'.txt', 'w')
              for index, value in df_BI_DEV.iterrows():
                  if m in str(value['BRAND_NAME']):
                      fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                      counter += 1
              print(m, ':', counter)
              fh.close()
```

```
Fill type word frequecy (brand name):

SALINE : 3711
SILICONE : 2325
GEL : 4884
COHESIVE : 160
```

```
In [103]: print('Surface type word frequecy (generic name):\n')
          for m in surface_type:
              counter = 0
              fh = open('WF/surface_type_GENERIC_NAME/'+m+'.txt', 'w')
              for index, value in df_BI_DEV.iterrows():
                  if m in str(value['GENERIC_NAME']):
                      fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                      counter += 1
              print(m, ':', counter)
              fh.close()
```

```
Surface type word frequecy (generic name):

SMOOTH : 164
TEXTURED : 379
BIOCELL : 8
```

```
In [104]: print('Fill type word frequecy (generic name):\n')
          for m in fill_type:
              counter = 0
              fh = open('WF/fill_type_GENERIC_NAME/'+m+'.txt', 'w')
              for index, value in df_BI_DEV.iterrows():
                  if m in str(value['GENERIC_NAME']):
                      fh.write(str(value['MDR_REPORT_KEY']) + '\n')
                      counter += 1
              print(m, ':', counter)
              fh.close()
```

```
Fill type word frequecy (generic name):

SALINE : 5217
SILICONE : 3704
GEL : 4932
COHESIVE : 28
```

```
In [ ]:
```